Математички Билтен
25 (LI)
2001 (73–90)
Скопје, Македонија

ISSN 0351–336X

SOME INEQUALITIES FOR TWO CSISZÁR DIVERGENCES AND APPLICATIONS

S. S. Dragomir

Abstract

Some inequalities for the Csiszár divergences of two mappings with applications to the variational distance, Kullback-Leibler distance, Hellinger discrimination, Chi-Square distance, Bhattacharyya distance, Jeffreys divergence, etc... are given.

1. Introduction

Given a convex function $f : [0,\infty) \to \mathbf{R}$, 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 $\mathbf{R}^n$. 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

$$f(0) = \lim_{t\to 0+} f(t), \qquad 0f\left(\frac{0}{0}\right) = 0,$$

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

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

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

Theorem 2. (*Jensen's inequality*) *Let $f : [0, \infty) \to \mathbf{R}$ be convex. Then for any $p, q \in \mathbf{R}_+^n$ 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

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

It is natural to consider the following corollary.

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

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

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

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

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

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

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

1. **Kullback-Leibler distance** ([12b]). 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 $f(t) = |t-1|$, $t \in \mathbf{R}_+$, 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** α**-order entropy** ([14]). The α- *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. The Results

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

Theorem 3. *Let $\Phi : \mathbf{R}_+ \to \mathbf{R}$ be differentiable convex. Then for all $p,q \in \mathbf{R}_+^n$ 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 $\Phi' : (0,\infty) \to \mathbf{R}$ 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 another 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 $\frac{p_1}{q_1} = \cdots = \frac{p_n}{q_n}$.

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 $f: C \subset \mathbf{R} \to \mathbf{R}$, where C is an interval (in [30], the definition was considered in general normed spaces), is

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

(ii) β- *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-upper convex.

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

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

(i) *If* Φ *is* α*-lower convex on* $\mathbf{R}_+$*, 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* β*-upper convex on* $\mathbf{R}_+$*, 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* $\mathbf{R}_+$*, 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.

We start with the following result.

Theorem 6. *Let* $f, g: [0, \infty) \to \mathbf{R}$ *be two mappings such that* $f(1) = g(1) = 0$. *If there exists the real constants* m, M *such that*

$$\begin{gathered} m\,|f(x) - f(y)| \le |g(x) - g(y)| \le M\,|f(x) - f(y)| \\ \text{for all} \quad x, y \in [r, R] \subset (0, \infty), \end{gathered} \tag{2.9}$$

then we have the inequality:

$$mI_{|f|}(p,q) \le I_{|g|}(p,q) \le MI_{|f|}(p,q) \tag{2.10}$$

for all p,q *probability distributions with* $0 < r \le \frac{p_i}{q_i} \le R < \infty$ *for all* $i \in \{1,...,n\}$.

Proof. By (2.9) it follows that

$$\begin{aligned} m\left|f\left(\frac{p_i}{q_i}\right)\right| &= m\left|f\left(\frac{p_i}{q_i}\right) - f(1)\right| \le \left|g\left(\frac{p_i}{q_i}\right) - g(1)\right| = \left|g\left(\frac{p_i}{q_i}\right)\right| \\ &\le M\left|f\left(\frac{p_i}{q_i}\right) - f(1)\right| = M\left|f\left(\frac{p_i}{q_i}\right)\right| \end{aligned} \tag{2.11}$$

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

If we multiply (2.11) by $q_i \ge 0$ and sum the obtained inequalities, we may deduce (2.10). □

Corollary 2. *Assume that the mappings* $f,g\colon [0,\infty) \to \mathbf{R}$ *are as above and* f,g *are differentiable on* (r,R) *with* $f'(t) \ne 0$ *for* $t \in (r,R)$ *and*

$$-\infty < m = \inf_{t\in(r,R)} \left|\frac{g'(t)}{f'(t)}\right|, \quad \sup_{t\in(r,R)} \left|\frac{g'(t)}{f'(t)}\right| = M < \infty, \tag{2.12}$$

then we have the inequality (2.10) *for all* p,q *as above.*

Proof. We use the following Cauchy's theorem:

If $\gamma, \phi\colon [a,b] \to \mathbf{R}$ are continuous and differentiable on (a,b) and $\phi'(t) \ne 0$ for all $t \in (a,b)$, then there exists a $c \in [a,b]$ such that

$$\frac{\gamma(b) - \gamma(a)}{\phi(b) - \phi(a)} = \frac{\gamma'(c)}{\phi'(c)}.$$

Now, suppose that $x,y \in [r,R]$ and $x < y$. Then, by Cauchy's theorem, we have

$$m \le \left|\frac{g(x) - g(y)}{f(x) - f(y)}\right| = \left|\frac{g'(z)}{f'(z)}\right| \le M$$

and then we can conclude that for any $x,y \in [r,R]$ we have

$$m\left|f(x) - f(y)\right| \le \left|g(x) - g(y)\right| \le M\left|f(x) - f(y)\right|.$$

Applying Theorem 6, we deduce (2.10). □

The following corollary for the variational distance holds.

Corollary 3. *Let* $g\colon [0,\infty) \to \mathbf{R}$ *be a mapping such that* $g(1) = 0$. *If there exists the real constants* n, N *such that*

$$n|x-y| \le |g(x) - g(y)| \le N|x-y| \quad \text{for all} \quad x, y \in [r, R], \tag{2.13}$$

then we have the inequality

$$nV(p,q) \le I_{|g|}(p,q) \le NV(p,q) \tag{2.14}$$

for any probability distribution p, q *with* $0 < r \le \frac{p_i}{q_i} \le R < \infty$ *for all* $i \in \{1, ..., n\}$.

The proof is obvious by Theorem 6, choosing $f(x) = x - 1$.

Corollary 4. *Assume that the mapping* g *is continuous on* $[a,b]$ *and differentiable on* (a,b) *and*

$$-\infty < n = \inf_{t\in(r,R)} |g'(t)|, \quad \sup_{t\in(r,R)} |g'(t)| = N < \infty.$$

Then we have the inequality (2.14) *for all* p, q *as above.*

3. Some Particular Cases in Terms of the Variational Distance

We start with the following result.

Proposition 1. *Let* $0 < r \le \dfrac{p_i}{q_i} \le R < \infty$ $(i = 1, ..., n)$. *Then we have the inequality*

$$0 \le KL(p,q) \le \begin{cases} [\ln R + 1]\, V(p,q) & \text{if} \quad r \ge e^{-1}, \\ \max\{\ln R + 1; |\ln R + 1|\}\, V(p,q) & \text{if} \quad r < e^{-1}. \end{cases} \tag{3.1}$$

Proof. Consider the mapping $g\colon (0,\infty) \to \mathbf{R}$, $g(t) = t\ln t$. Then $g'(t) = \ln t + 1$ and obviously,

$$M := \sup_{t\in(r,R)} |g'(t)| = \begin{cases} \ln R + 1 & \text{if} \quad r \ge e^{-1}, \\ \max\{\ln R + 1; |\ln R + 1|\} & \text{if} \quad r < e^{-1}. \end{cases}$$

Applying Corollary 4, we can state

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

By the generalised triangle inequality, we have

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

and the inequality (3.1) is proved. □

Let us introduce the *modified Kullback-Leibler* distance

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

Then obviously,

$$K(p,q)\leq|KL|(p,q)\quad \text{forall}\quad p,q\in\mathbf{R}^n. \tag{3.2}$$

For this modified distance, we may prove the following as well.

Proposition 2. *Let* $0<r\leq\frac{p_i}{q_i}\leq R<\infty$ $(i=1,\ldots,n)$. *Then we have the inequality*

$$(\ln r+1)V(p,q)\leq|KL|(p,q)\leq(\ln R+1)V(p,q), \tag{3.3}$$

provided that $r\geq e^{-1}$.

Proof. The second inequality in (2.11) has been proven above.

For the first inequality, we can apply Corollary 4 by observing that for $g(t)=t\ln t$, and $r\geq e^{-1}$,

$$\inf_{t\in[r,R]}|g'(t)|=\ln r+1.$$

We omit the details. □

The following proposition also holds.

Proposition 3. *Let* $0<r\leq\frac{p_i}{q_i}\leq R<\infty$ $(i=1,\ldots,n)$. *Then we have the inequality:*

$$KL(q,p)\leq\frac{1}{r}V(p,q). \tag{3.4}$$

Proof. Consider the mapping $g:(0,\infty)\to\mathbf{R}$, $g(t)=\ln t$. Then $g'(t)=\frac{1}{t}$ and obviously,

$$M:=\sup_{t\in[r,R]}|g'(t)|=\frac{1}{r}.$$

Applying Corollary 3, we can state:

$$\sum_{i=1}^{n} q_i \left| \ln\left(\frac{p_i}{q_i}\right) \right| \leq \frac{1}{r} V(p,q) .$$

By the generalised triangle inequality, we have

$$K(q,p) = \sum_{i=1}^{n} q_i \ln\left(\frac{q_i}{p_i}\right) = \left| \sum_{i=1}^{n} q_i \ln\left(\frac{q_i}{p_i}\right) \right| \leq \sum_{i=1}^{n} q_i \left| \ln\left(\frac{p_i}{q_i}\right) \right| \leq \frac{1}{r} V(p,q)$$

and the proposition is proved. □

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

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

$$\frac{1}{R} V(p,q) \leq |KL|(q,p) \leq \frac{1}{r} V(p,q) . \tag{3.5}$$

Proof. The second inequality in (3.5) has been proven above. The first inequality follows by the first inequality in Corollary 4 by taking into account that

$$m = \inf_{t \in (r,R)} |g'(t)| = \frac{1}{R} .$$

Now, the following result for *Hellinger discrimination* holds.

Proposition 5. *Let $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ $(i = 1, ..., n)$. Then we have the inequality:*

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

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

$$g'(t) = \frac{1}{2} \cdot \frac{\sqrt{t} - 1}{\sqrt{t}}, \quad t \in (0, \infty)$$

and

$$n=\inf_{t\in[r,R]}|g'(t)|=\min\{|g'(r)|,|g'(R)|\}$$
$$=\frac{|g'(r)|+|g'(R)|-||g'(r)|-|g'(R)||}{2}$$
$$=\frac{\sqrt{R}-\sqrt{r}}{4\sqrt{rR}}-\left|\frac{\sqrt{R}+\sqrt{r}}{4\sqrt{rR}}-\frac{1}{2}\right|$$

and

$$N=\sup_{t\in[r,R]}|g'(t)|=\max\{|g'(r)|,|g'(R)|\}$$
$$=\frac{|g'(r)|+|g'(R)|+||g'(r)|-|g'(R)||}{2}$$
$$=\frac{\sqrt{R}-\sqrt{r}}{4\sqrt{rR}}+\left|\frac{\sqrt{R}+\sqrt{r}}{4\sqrt{rR}}-\frac{1}{2}\right|$$

respectively.

As $g(t)\geq 0$, then obviously

$$I_{|g|}(p,q)=I_g(p,q)=h^2(p,q).$$

Using (2.14), we obtain (3.6). □

Remark 1. *The inequality* (3.6) *is equivalent to*

$$\left|h^2(p,q)-\frac{\sqrt{R}-\sqrt{r}}{4\sqrt{rR}}V(p,q)\right|\leq\left|\frac{\sqrt{R}+\sqrt{r}}{4\sqrt{rR}}-\frac{1}{2}\right|V(p,q) \qquad (3.7)$$

Now, we point out some inequalities for the *chi-square distance.*

Proposition 6. *Let* $0<r\leq\frac{p_i}{q_i}\leq R<\infty$ $(i=1,...,n)$. *Then we have the inequality*

$$[R-r-|R+r-2|]\,V(p,q)\leq D_{\chi^2}(p,q)\leq[R-r+|R+r-2|]\,V(p,q). \qquad (3.8)$$

Proof. Consider the mapping $g:(0,\infty)\to\mathbf{R}$, $g(t)=(t-1)^2$. Then obviously $g'(t)=2(t-1)$ and

$$n=\inf_{t\in[r,R]}|g'(t)|=\min\{|g'(r)|,|g'(R)|\}=R-r-|R+r-2|$$

and

$$N=R-r+|R+r-2|.$$

Using the inequality (2.14), and taking into account that $g(t) \geq 0$, $t \in \mathbf{R}$, and

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

we deduce (3.8). □

Remark 2. *The inequality* (3.8) *is equivalent with*

$$\left| D_{\chi^2}(p,q) - (R-r) V(p,q) \right| \leq |R + r - 2| V(p,q) . \tag{3.9}$$

We point out now some inequalities for the Bhattacharyya distance.

Proposition 7. *Let* $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ $(i = 1, ..., n)$. *Then we have the inequality:*

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

Proof. Consider the mapping $g(t) = 1 - \sqrt{t}$, $t \in (0, \infty)$. Then $g(1) = 0$, $g'(t) = -\dfrac{1}{2\sqrt{t}}$ and

$$N = \sup_{t \in [r,R]} |g'(t)| = \sup_{t \in [r,R]} \frac{1}{2\sqrt{t}} = \frac{1}{2\sqrt{r}} .$$

Applying Corollary 4, we may state

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

which is equivalent to

$$\sum_{i=1}^{n} |q_i - \sqrt{p_i q_i}| \leq \frac{1}{2\sqrt{r}} V(p,q) . \tag{3.11}$$

Using the generalised triangle inequality, we obtain

$$\sum_{i=1}^{n} |q_i - \sqrt{p_i q_i}| \geq \left| \sum_{i=1}^{n} (q_i - \sqrt{p_i q_i}) \right| = |1 - B(p,q)| = 1 - B(p,q) . \quad \square$$

If we define the following distance $\tilde{B}(p,q) := \sum_{i=1}^{n} \sqrt{q_i} \left| \sqrt{q_i} - \sqrt{p_i} \right|$, then we may state the following proposition as well.

Proposition 8. *Assume that p_i, q_i, r, R are as above. Then*

$$\frac{1}{2\sqrt{R}} V(p,q) \le \tilde{B}(p,q) \le \frac{1}{2\sqrt{r}} V(p,q). \tag{3.12}$$

The proof is obvious by Corollary 3 applied for the mapping $g(t) = 1 - \sqrt{t}$.

Now, let us consider the *harmonic distance*

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

The following proposition holds.

Proposition 9. *Assume that p_i, q_i, r, R are as above. Then we have the inequality:*

$$0 \le 1 - M(p,q) \le \frac{2}{(r+1)^2} V(p,q). \tag{3.13}$$

Proof. Consider the mapping $g(t) = 1 - \frac{2t}{t+1}$. Then $g(1) = 0$, $g'(t) = -\frac{2}{(t+1)^2}$ and

$$N := \sup_{t \in [r,R]} |g'(t)| = \frac{2}{(r+1)^2}.$$

Applying Corollary 4, we can state that

$$\sum_{i=1}^{n} q_i \left| 1 - \frac{2\frac{p_i}{q_i}}{\frac{p_i}{q_i} + 1} \right| \le \frac{2}{(r+1)^2} V(p,q),$$

which is clearly equivalent to:

$$\sum_{i=1}^{n} \frac{q_i |p_i - q_i|}{p_i + q_i} \le \frac{2}{(r+1)^2} V(p,q). \tag{3.14}$$

Using the generalised triangle inequality, we get (3.13). □

If we introduce the divergence measure:

$$\tilde{M}(p,q) := \sum_{i=1}^{n} q_i \cdot \frac{|p_i - q_i|}{p_i + q_i} = I_l(p,q),$$

where $l(t) = \dfrac{|t-1|}{t+1}$, $t > 0$, then we have the following proposition.

Proposition 10. *With the above assumptions, we have*

$$\frac{2}{(R+1)^2} V(p,q) \le \tilde{M}(p,q) \le \frac{2}{(r+1)^2} V(p,q). \tag{3.15}$$

Finally, let us consider the *Jeffreys distance*

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

The following proposition holds.

Proposition 11. *Assume that* $0 < r \le \dfrac{p_i}{q_i} \le R < \infty$. *Then we have the inequality:*

$$\begin{aligned}
&\left[\frac{R-r}{2rR} + \ln\sqrt{\frac{R}{r}} - \left|\frac{R-r}{2rR} - \ln\sqrt{rR} - 1\right|\right] V(p,q) \le J(p,q) \\
&\le \left[\frac{R-r}{2rR} + \ln\sqrt{\frac{R}{r}} + \left|\frac{R-r}{2rR} - \ln\sqrt{rR} - 1\right|\right] V(p,q).
\end{aligned} \tag{3.16}$$

Proof. Consider the mapping $g(t) = (t-1)\ln t$, $t > 0$. Then, obviously $g'(t) = \ln t - \dfrac{1}{t} + 1$, $g''(t) = \dfrac{t+1}{t^2}$, which shows that $g'(\cdot)$ is strictly increasing on $(0,\infty)$ and $g'(1) = 0$. Then

$$n = \inf_{t\in[r,R]} |g'(t)| = \min\{|g'(r)|, |g'(R)|\} = \frac{R-r}{2rR} + \ln\sqrt{\frac{R}{r}} - \left|\frac{R+r}{2rR} - \ln\sqrt{rR} - 1\right|$$

and

$$N = \sup_{t\in[r,R]} |g'(t)| = \max\{|g'(r)|, |g'(R)|\} = \frac{R-r}{2rR} + \ln\sqrt{\frac{R}{r}} + \left|\frac{R+r}{2rR} - \ln\sqrt{rR} - 1\right|.$$

In addition, as

$$\begin{aligned}
I_{|g|}(p,q) &= \sum_{i=1}^{n} q_i \left|\left(\frac{p_i}{q_i} - 1\right)\right| \left|\ln\left(\frac{p_i}{q_i}\right)\right| = \sum_{i=1}^{n} |p_i - q_i| \, |\ln p_i - \ln q_i| \\
&= \sum_{i=1}^{n} (p_i - q_i)(\ln p_i - \ln q_i) = I(p,q),
\end{aligned}$$

then by (2.14), we deduce (3.16).

Remark 3. *The above inequality* (3.16) *is equivalent to*

$$\left|J(p,q)-\left[\frac{R-r}{2rR}+\ln\sqrt{\frac{R}{r}}\right]V(p,q)\right|\leq\left|\frac{R+r}{2rR}-\ln\sqrt{rR}-1\right|V(p,q). \quad (3.17)$$

4. Other Particular Cases

Let us consider the modified Kullback-Leibler divergence

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

where $p,q\in\mathbf{R}^n$.

We point out some estimates in terms of $|KL|$.

Proposition 12. *Assume that* $0<r\leq\frac{p_i}{q_i}\leq R<\infty$ $(i=1,...,n)$. *Then we have the inequality*

$$0\leq KL(p,q)\leq\left[\frac{R-r}{2}+\ln\sqrt{\frac{R^R}{r^r}}+\left|\frac{r+R}{2}+\ln\sqrt{R^Rr^r}\right|\right]|KL|(q,p). \quad (4.1)$$

Proof. Consider the mappings $g(t)=t\ln t$, $f(t)=\ln t$, $t>0$. Then $h(t):=\frac{g'(t)}{f'(t)}=t\ln t+t.$

We observe that

$$\begin{aligned} M &= \sup_{t\in[r,R]}|h(t)|=\max\{|h(r)|,|h(R)|\} \\ &= \frac{R-r}{2}+\ln\sqrt{\frac{R^R}{r^r}}+\left|\frac{r+R}{2}+\ln\sqrt{R^Rr^r}\right|. \end{aligned}$$

Applying Corollary 2, we may write

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

and as, by the generalised triangle inequality, we have

$$\sum_{i=1}^{n}p_i\left|\ln\left(\frac{p_i}{q_i}\right)\right|\geq|KL(p,q)|=KL(p,q)\geq0,$$

the inequality (4.1) is proved. □

We now compare the Hellinger discrimination with $|KL|$.

Proposition 13. *Let p_i, q_i, r, R be as in Proposition 12. Then we have the inequality:*

$$\begin{aligned}\frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}-\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right|\right]|KL|(q,p)\le h^2(p,q)\\ \le\frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}+\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right|\right]|KL|(q,p).\end{aligned}\tag{4.2}$$

Proof. Consider the mappings $g(t)=\frac{1}{2}\left(\sqrt{t}-1\right)^2$, $f(t)=\ln t$, $t>0$. Then

$$h(t):=\frac{g'(t)}{f'(t)}=\frac{1}{2}\left(\frac{\sqrt{t}-1}{\sqrt{t}}\right)\cdot t=\frac{1}{2}\left(\sqrt{t}-1\right)\sqrt{t},\qquad t>0.$$

We observe that

$$\begin{aligned}m&=\inf_{t\in[r,R]}|h(t)|=\min\{|h(r)|,|h(R)|\}\\&=\frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}-\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right|\right]\end{aligned}$$

and, analogously,

$$\begin{aligned}M&=\sup_{t\in[r,R]}|h(t)|=\max\{|h(r)|,|h(R)|\}\\&=\frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}+\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right|\right].\end{aligned}$$

Now, as $g(t)\ge 0$, we have

$$I_{|g|}(p,q)=I_g(p,q)=h^2(p,q)$$

and then, by Corollary 2, we deduce (4.2). □

Remark 4. *The above inequality is equivalent with*

$$\left|h^2(p,q)-\frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}\right]|KL(q,p)|\right|\le\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right||KL|(q,p).\tag{4.3}$$

We now compare the Chi-square distance with $|KL|$. The following proposition holds.

Proposition 14. *Let p_i, q_i, r, R be as above. Then*

$$\begin{aligned}&\left[(R-r)(R+r-1)-\left|R+r-\left(R^2+r^2\right)\right|\right]|KL|(q,p)\leq D_{\chi^2}(p,q)\\&\leq\left[(R-r)(R+r-1)+\left|R+r-\left(R^2+r^2\right)\right|\right]|KL|(q,p).\end{aligned}\tag{4.4}$$

Proof. Consider the mappings $g(t)=(t-1)^2$, $f(t)=\ln t$, $t>0$. Then $h(t)=\frac{g'(t)}{f'(t)}=2t(t-1)$.

We observe that

$$\begin{aligned}m=\inf_{t\in[r,R]}|h(t)|&=\frac{1}{2}\left[2r(1-r)+2R(R-1)-|2r(1-r)-2R(R-1)|\right]\\&=\left[r-r^2+R^2-R-\left|r-r^2-R^2+R\right|\right]\\&=R^2-r^2-(R-r)-\left|R+r-\left(R^2+r^2\right)\right|\\&=(R-r)(R+r-1)-\left|R+r-\left(R^2+r^2\right)\right|\end{aligned}$$

and

$$M=\sup_{t\in[r,R]}(h(t))=(R-r)(R+r-1)+\left|R+r-\left(R^2+r^2\right)\right|.$$

Now, as $g(t)\geq 0$, we have

$$I_{|g|}(p,q)=I_g(p,q)=D_{\chi^2}(p,q)$$

and then, by Corollary 2, we deduce (4.4). □

Remark 5. *The above inequality is equivalent with*

$$\left|D_{\chi^2}(p,q)-(R-r)(R+r-1)|KL|(q,p)\right|\leq\left|R+r-\left(R^2+r^2\right)\right||KL|(q,p).\tag{4.5}$$

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 Φ-divergence in information theory, *submitted.*

НЕКОИ НЕРАВЕНСТВА ЗА ДВЕ ЧЕЗАРОВИ ДИВЕРГЕНЦИИ И ПРИМЕНА

Р е з и м е

Дадени се некои неравенства за Чезарови дивергенции од две пресликувања со примена на варијационо растојание, Kullback-Leibler-ово растојание, Hellingen-ово растојание, Chi-Square-ово растојание, Bhattacharyya-ово растојание, Jeffreys-ово растојание итн.

School of Communications and Informatics

Victoria University of Technology

PO Box 14428;

Victoria; 8001; AUSTRALIA

e-mail: sever.dragomir@vu.edu.au
URL:http://rgmia.vu.edu.au/SSDragomirWeb.html