Sever Silvestru Dragomir

Bounds for f-divergences under likelihood ratio constraints

Applications of Mathematics, Vol. 48 (2003), No. 3, 205–223

Persistent URL: http://dml.cz/dmlcz/134528

BOUNDS FOR f-DIVERGENCES UNDER LIKELIHOOD RATIO CONSTRAINTS

S. S. Dragomir, Melbourne City

(Received December 10, 1999, in revised version June 7, 2001)

Abstract. In this paper we establish an upper and a lower bound for the f-divergence of two discrete random variables under likelihood ratio constraints in terms of the Kullback-Leibler distance. Some particular cases for Hellinger and triangular discrimination, χ^2-distance and Rényi's divergences, etc. are also considered.

Keywords: f-divergence, divergence measures in information theory, Jensen's inequality, Hellinger and triangular discrimination

MSC 2000: 94A17, 26D15

1. Introduction

Given a convex function $f\colon [0,\infty) \to \mathbb{R}$, 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

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

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

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

Proposition 2 (Jensen's inequality). *Let $f\colon [0,\infty) \to \mathbb{R}$ 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*

$$I_f(p,q) \geqslant Q_n f\Big(\frac{P_n}{Q_n}\Big). \tag{1.2}$$

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

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

It is natural to consider the following corollary.

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

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

Then for any $p,q \in [0,\infty)^n$ with $P_n = Q_n$ we have the inequality

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

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

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

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

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

We now give some examples of divergence measures in information theory which are particular cases of f-divergences.

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

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

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 $f(t) = |t-1|$, $t \in [0,\infty)$, then we have

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

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

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

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

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

(4) **Triangular discrimination** ([24]). We define the *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 $f(t) = (t-1)^2/(t+1)$, $t \in (0,\infty)$, then

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

Note that $\sqrt{\Delta(p,q)}$ is known in literature as the Le Cam distance.

(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's divergences** ([14]). For $\alpha \in \mathbb{R} \setminus \{0, 1\}$, consider

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

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

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

Rényi's divergences $R_\alpha(p, q) = \alpha^{-1}(\alpha - 1)^{-1} \ln[\varrho_\alpha(p, q)]$ have been introduced for all real orders $\alpha \neq 0$, $\alpha \neq 1$ (and continuously extended for $\alpha = 0$ and $\alpha = 1$) in [31], where the reader may find many inequalities valid for these divergences, without, as well as with, some restrictions for p and q.

For other examples of divergence measures, see the paper [22] and the books [31] and [32], where further references are given.

2. Some inequalities between the f-divergence and the Kullback-Leibler distance

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

Proposition 3. *Let $\Phi\colon [0, \infty) \to \mathbb{R}$ be differentiable and convex. Then for all $p, q \in [0, \infty)^n$ we have the inequality*

$$\Phi'(1)(P_n - Q_n) \leqslant I_\Phi(p, q) - Q_n \Phi(1) \leqslant I_{\Phi'}\Big(\frac{p^2}{q}, p\Big) - I_{\Phi'}(p, q), \tag{2.1}$$

where $P_n = \sum\limits_{i=1}^{n} p_i > 0$, $Q_n = \sum\limits_{i=1}^{n} q_i > 0$, $\Phi'\colon (0, \infty) \to \mathbb{R}$ is the derivative of Φ, and $I_{\Phi'}\big(p^2/q, p\big) = \sum\limits_{i=1}^{n} p_i \Phi'\big(p_i/q_i\big)$.

If Φ is strictly convex and p_i, $q_i > 0$ $(i = 1, \ldots, n)$, then equality holds in (2.1) iff $p = q$.

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

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

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

A result similar to the above theorem has been presented in another paper by the author [29].

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

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

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

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

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

(i) *α-lower convex on C* if $f(t) - \frac{1}{2}\alpha t^2$ is convex on C;

(ii) *β-upper convex on C* if $\frac{1}{2}\beta t^2 - f(t)$ is convex on C;

(iii) *(α, β)-convex on C* (with $\alpha \leqslant \beta$) if it is both α-lower convex and β-upper convex.

In [30], among other, the author has proved the following result for the f-divergence.

Proposition 5. *Let $\Phi\colon [0,\infty) \to \mathbb{R}$ and $p, q \in [0,\infty)^n$ with $P_n = Q_n$.*

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

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

(ii) *If Φ is β-upper convex on $[0,\infty)$, then we have the inequality*

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

(iii) *If Φ is (α,β)-convex on $[0,\infty)$, then we have the sandwich inequality*

$$\frac{\alpha}{2} D_{\chi^2}(p,q) \leqslant I_\Phi(p,q) - Q_n \Phi(1) \leqslant \frac{\beta}{2} D_{\chi^2}(p,q), \tag{2.6}$$

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 simpler inequalities

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

and

$$\frac{\alpha}{2} D_{\chi^2}(p,q) \leqslant I_\Phi(p,q) \leqslant \frac{\beta}{2} D_{\chi^2}(p,q). \tag{2.8}$$

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

The following result concerning an upper and a lower bound for the f-divergence in terms of the Kullback-Leibler distance $D(p,q)$ holds. This result complements, in a sense, the results presented above in Proposition 5.

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

(i) *f is twice differentiable on (r,R), where $0 \leqslant r \leqslant 1 \leqslant R \leqslant \infty$;*

(ii) *there exist constants m, M such that*

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

If p, q are discrete probability distributions satisfying the assumption

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

then we have the inequality

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

P r o o f. Define a mapping $F_m\colon (0,\infty) \to \mathbb{R}$, $F_m(t) = f(t) - mt\ln t$. Then $F_m(\cdot)$ is normalised, twice differentiable and since

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

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

$$0 \leqslant I_{F_m}(p,q) = I_f(p,q) - mI_{(\cdot)\ln(\cdot)}(p,q) = I_f(p,q) - mD(p,q) \tag{2.13}$$

from where the first inequality in (2.11) results.

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

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

R e m a r k 2. It is important to note that if f is twice differentiable on $(0,\infty)$ and $0 < m \leqslant tf''(t) \leqslant M < \infty$ for any $t \in (0,\infty)$, then inequality (2.11) holds for any probability distributions p, q.

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

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

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

then

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

and

$$\begin{aligned} &m[D(\lambda p^{(1)} + (1-\lambda)p^{(2)}, \lambda q^{(1)} + (1-\lambda)q^{(2)}) \\ &\qquad - \lambda D(p^{(1)}, q^{(1)}) - (1-\lambda)D(p^{(2)}, q^{(2)})] \\ &\quad \leqslant I_f(\lambda p^{(1)} + (1-\lambda)p^{(2)}, \lambda q^{(1)} + (1-\lambda)q^{(2)}) \\ &\qquad - \lambda I_f(p^{(1)}, q^{(1)}) - (1-\lambda)I_f(p^{(2)}, q^{(2)}) \\ &\quad \leqslant M[D(\lambda p^{(1)} + (1-\lambda)p^{(2)}, \lambda q^{(1)} + (1-\lambda)q^{(2)}) \\ &\qquad - \lambda D(p^{(1)}, q^{(1)}) - (1-\lambda)D(p^{(2)}, q^{(2)})] \end{aligned} \tag{2.16}$$

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

P r o o f. By (2.14) we have

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

and

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

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

It is already known that the mappings F_m, F_M as defined in Theorem 1 are convex and normalised.

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

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

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

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

We omit the details. □

R e m a r k 3. If $m > 0$ in (2.9), then the inequality (2.11) is a better result than the positivity property of the f-divergence. The same will apply to the joint convexity of the f-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 3. *Let $f\colon [0,\infty)\to\mathbb{R}$ be a normalised mapping, i.e., $f(1)=0$, and satisfy the assumptions*

(i) *f is twice differentiable on (r,R), where $0\leqslant r\leqslant 1\leqslant R\leqslant\infty$;*

(ii) *there exist constants m, M such that*

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

If p, q are discrete probability distributions satisfying the assumption

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

then we have the inequality

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

P r o o f. We know (see the proof of Theorem 1) that the mapping $F_m\colon (0,\infty)\to\mathbb{R}$, $F_m(t)=f(t)-mt\ln t$ is normalised, twice differentiable and convex on (r,R).

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

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

However,

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

and

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

Consequently, by (2.23) we have

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

As simple computation shows that $D\big(p,p^2/q\big) = -D(p,q)$, 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) to F_M, we may write

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

However,

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

and hence, by (2.24), we get

$$MD(p,q) - I_f(p,q) \leqslant -MD\Big(p,\frac{p^2}{q}\Big) + M - I_{f'}\Big(\frac{p^2}{q},p\Big) + MD(q,p) - M + I_{f'}(p,q),$$

which is equivalent to the first part of (2.22). □

R e m a r k 4. The inequality (2.22) is obviously equivalent to the following one:

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

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 1, we are able to point out the following particular cases which may be of interest in information theory.

Proposition 6. *Let p, q be two probability distributions with the property that*

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

Then we have the inequality

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

P r o o f. Consider the mapping $f\colon [r,R] \to \mathbb{R}$, $f(t) = -\ln t$. Define $g(t) = tf''(t) = t \cdot \big(1/t^2\big) = 1/t$. Then obviously

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

Also,

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

Now, using (2.11) with $m = 1/R$ and $M = 1/r$, we deduce the desired inequality. □

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

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

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

$$\Big|\frac{p_i}{q_i} - 1\Big| \leqslant \varepsilon \quad \text{for all } i \in \{1, \ldots, n\}. \tag{3.4}$$

Then we have the inequality

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

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

Proposition 7. *Let p, q be two probability distributions satisfying the condition* (3.1). *Then we have the inequality*

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

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

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

Since

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

we deduce the desired inequality by applying (2.11) for $m = 2r$ and $M = 2R$. □

Remark 5. The following inequality is well known in literature:

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

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

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

We remark that if $\frac{1}{2r} \leqslant 1$, i.e., $r \geqslant \frac{1}{2}$, 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*

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

The following inequality connecting the Kullback-Leibler distance with $h(p,q)$, defined in Introduction, holds.

Proposition 8. *Assume that the probability distributions* p,q *satisfy the condition* (3.1). *Then we have the inequality*

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

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

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

Then obviously

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

Since

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

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

Remark 6. The following inequality is well known in literature (see for example [25]):

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

for any probability distributions p,q.

From the second inequality in (3.9) we have

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

We remark that if $4\sqrt{r} \geqslant 2$, i.e., $r \geqslant \frac{1}{4}$, 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 9. *Assume that the probability distributions* p,q *satisfy the condition* (3.1).

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

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

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

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

P r o o f. 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

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

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 two cases:

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

$$\sup_{t\in[r,R]} g(t) = \frac{32}{27}$$

and

$$\inf_{t\in[r,R]} g(t) = \min[g(r), g(R)] = \min\left\{\frac{8r}{(r+1)^3}, \frac{8R}{(R+1)^3}\right\}.$$

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

$$\sup_{t\in[r,R]} g(t) = g(r) = \frac{8r}{(r+1)^3}$$

and

$$\inf_{t\in[r,R]} g(t) = g(R) = \frac{8R}{(R+1)^3}.$$

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

R e m a r k 7. It is clear, by the above arguments, that for every probability distribution we have the inequality

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

We know (see Topsoe [24]) that

$$2h^2(p,q) \leqslant \Delta(p,q) \leqslant 4h^2(p,q). \tag{3.15}$$

Now, as $D(p,q) \geqslant 2h^2(p,q)$, we obtain

$$\Delta(p,q) \leqslant 2D(p,q), \tag{3.16}$$

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

Let us compare the *Rényi α-divergence* with the Kullback-Leibler distance. The following proposition holds:

Proposition 10. *Assume that probability distributions* p, q *satisfy the condition* (3.1). *Then*

$$\begin{aligned} \alpha(\alpha-1)r^{\alpha-1}D(p,q)+1 &\leqslant \exp[\alpha(\alpha-1)R_\alpha(p,q)] \\ &\leqslant \alpha(\alpha-1)R^{\alpha-1}D(p,q)+1 \end{aligned} \tag{3.17}$$

for $\alpha > 1$.

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

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

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

$$\alpha(\alpha-1)r^{\alpha-1}D(p,q) \leqslant \sum_{i=1}^n q_i\Big[\Big(\frac{p_i}{q_i}\Big)^\alpha - 1\Big] \leqslant \alpha(\alpha-1)R^{\alpha-1}D(p,q),$$

i.e.,

$$\alpha(\alpha-1)r^{\alpha-1}D(p,q)+1 \leqslant \varrho_\alpha(p,q) \leqslant \alpha(\alpha-1)R^{\alpha-1}D(p,q)+1$$

and the proposition is proved. □

We define the *Bhattacharyya distance* (see [27]) by $B(p,q)=-\ln[\gamma(p,q)]$, where

$$\gamma(p,q)=\sum_{i=1}^{n}\sqrt{p_iq_i}.$$

The following proposition holds.

Proposition 11. *Assume that the probability distributions* p,q *satisfy the condition* (3.1). *Then*

$$4\sqrt{r}[1-\exp[-B(p,q)]]\leqslant D(p,q)\leqslant 4\sqrt{R}[1-\exp[-B(p,q)]]. \tag{3.18}$$

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

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

Applying the inequality (2.11), we have

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

i.e.,

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

which is equivalent to (3.18). □

We define the *harmonic divergence* by $M(p,q)=1-m(p,q)$, where

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

The following proposition holds:

Proposition 12. *Assume that* p,q *are two discrete probability distributions. Then*

$$0\leqslant M(p,q)\leqslant\frac{16}{27}D(p,q). \tag{3.19}$$

P r o o f. Consider the mapping $f\colon (0,\infty)\to\mathbb{R}$, $f(t)=2t/(t+1)-1$. Then f is normalised,

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

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

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

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

$$\inf_{t\in(0,\infty)} g(t)=g\Big(\frac{1}{2}\Big)=-\frac{16}{27},$$
$$\sup_{t\in(0,\infty)} g(t)=0.$$

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

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

which is equivalent to

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

and the inequality (3.19) is proved. □

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

Proposition 13. *Assume that* p,q *satisfy the condition* (2.10).
(i) *If* $r\in\big(0,\frac{1}{2}\big)$, *then*

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

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

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

P r o o f.

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

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

and, applying (2.11), we may write

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

and the inequality (3.20) is proved.

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

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

that is,

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

Applying (2.11), we deduce (3.21). □

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

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

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

The following proposition also holds.

Proposition 14. *Assume that p, q satisfy the condition* (2.10). *Then*

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

P r o o f. Consider $f(t) = (t-1)\ln t$. Then $f'(t) = \ln t - 1/t + 1$ and $f''(t) = (t+1)/t^2$. Define $g(t) = tf''(t) = 1 + 1/t$. 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 $M = (r+1)/r$, $m = (R+1)/R$, we obtain the desired result. □

R e m a r k 8. Similar results can be obtained by applying Theorem 3, but we omit the details.

Acknowledgement. The author would like to thank the anonymous referee for some valuable comments and for letting us know about references [31] and [32].

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, pp. 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, J. Körner*: Information Theory: Coding Theorem for Discrete Memory-less Systems. Academic Press, New York, 1981.

[4] Maximum Entropy and Bayesian Methods in Applied Statistics (J. H. Justice, ed.). 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, pp. 1–44.

[6] *I. Burbea, 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, ed.). Springer-Verlag, Berlin, 1975.

[10] *R. M. Leahy, C. E. Goutis*: An optimal technique for constraint-based image restoration and mensuration. IEEE Trans. on Acoustics, Speech and Signal Processing *34* (1986), 1629–1642.

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

[12] *S. Kullback, R. A. Leibler*: On information and sufficiency. Ann. Math. Statistics *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, N. M. Ionescu*: Some converse of Jensen's inequality and applications. Anal. Numer. Theor. Approx. *23* (1994), 71–78.

[16] *S. S. Dragomir, C. J. Goh*: A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory. Math. Comput. Modelling *24* (1996), 1–11.

[17] *S. S. Dragomir, C. J. Goh*: Some counterpart inequalities in for a functional associated with Jensen's inequality. J. Inequal. Appl. *1* (1997), 311–325.

[18] *S. S. Dragomir, C. J. Goh*: Some bounds on entropy measures in information theory. Appl. Math. Lett. *10* (1997), 23–28.

[19] *S. S. Dragomir, 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. Ph.D. Thesis. Univ. of Zagreb, 1999. (In Croatian.)
[21] *S. S. Dragomir, J. Unde and M. Scholz*: Some upper bounds for relative entropy and applications. Comput. Math. Appl. *39* (2000), 257–266.
[22] *J. N. Kapur*: A comparative assessment of various measures of directed divergence. Advances Manag. Stud. *3* (1984), 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* (1999), 85–98.
[25] *D. Dacunha-Castelle*: Vitesse de convergence pour certains problemes statistiques. In: Ecole d'été de Probabilités de Saint-Flour, VII (Saint-Flour, 1977). Lecture Notes in Math. Vol. 678. Springer, Berlin, 1978, pp. 1–172.
[26] *H. Jeffreys*: An invariant form for the prior probability in estimation problems. Proc. Roy. Soc. London Ser. A *186* (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. Math. J. Ibaraki Univ. (Japan) *33* (2001), 35–50.
[31] *F. Liese, I. Vajda*: Convex Statistical Distances. Teubner Verlag, Leipzig, 1987.
[32] *I. Vajda*: Theory of Statistical Inference and Information. Kluwer, Boston, 1989.

Author's address: *S. S. Dragomir*, School of Communications and Informatics, Victoria University of Technology, P.O. Box 14428, Melbourne City MC, Victoria 8001, Australia, e-mail: `sever.dragomir@vu.edu.au`.