APPROXIMATIONS FOR CSISZÁR f–DIVERGENCE VIA MIDPOINT INEQUALITIES

S. S. DRAGOMIR, V. GLUŠČEVIĆ, AND C. E. M. PEARCE

Abstract. Using two midpoint inequalities arising in numerical integration we derive, for an absolutely continuous function f, an approximation for the Csiszár f–divergence of two probability distributions over a finite alphabet. Applications are made to some common divergence measures such as Kullback–Leibler distance, Hellinger discrimination and Renyi α–entropy.

1. Introduction

A common situation in information theory is the following. Two probability distributions $p = (p_1, \ldots, p_n)$, $q = (q_1, \ldots, q_n)$ are defined over an alphabet $\{a_i | 1 \le i \le n\}$, p_i, q_i being the point probabilities associated with event a_i $(i = 1, \ldots, n)$. For example, p, q might represent *a priori* and *a posteriori* probability distributions associated with the alphabet.

It is useful to be able to quantify in some way the difference between such distributions p, q. A number of ways have been suggested for doing this. Thus the *variational distance* (l_1-distance) and *information divergence* (Kullback–Leibler divergence) are defined respectively by

$$V(p,q) := \sum_{i=1}^{n} |p_i - q_i|, \tag{1.1}$$

$$D(p,q) := \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}. \tag{1.2}$$

(see for example [10]). As in [1], we define the *triangular discrimination* between p and q by

$$\Delta(p,q) := \sum_{i=1}^{n} \frac{|p_i - q_i|^2}{p_i + q_i}. \tag{1.3}$$

The *Hellinger discrimination* h^2 is given by

$$h^2(p,q) := \frac{1}{2} \sum_{i=1}^{n} \left(\sqrt{p_i} - \sqrt{q_i}\right)^2. \tag{1.4}$$

For applications it is important to know how these divergences compare with one another. The basic relations between V, Δ and h^2 are

$$\frac{1}{2} V^2(p,q) \le \Delta(p,q) \le V(p,q)$$

1991 *Mathematics Subject Classification.* 94A17, 26D15.

Key words and phrases. Csiszár f–divergence, midpoint inequalities, divergence measures.

and

$$2h^2(p,q) \le \Delta(p,q) \le 4h^2(p,q)$$

(see LeCam [5] and Dacunha–Castelle [6]). From these we may deduce that

$$\frac{1}{8}V^2(p,q) \le h^2(p,q) \le \frac{1}{2}V(p,q).$$

The coefficients in these inequalities are best possible (*cf.* [1]).

The first half of this result has been improved by Kraft [7], who showed that

$$\frac{1}{8}V^2(p,q) \le h^2(p,q)\left(1 - \frac{1}{2}h^2(p,q)\right).$$

We note also the important inequality

$$D(p,q) \ge -2\ln\left(1 - h^2(p,q)\right)$$

(see Dacunha–Castelle [6]). It follows from this that

$$D(p,q) \ge 2h^2(p,q).$$

Again the coefficient 2 is best–possible (see [1]).

Csiszár [2]–[4] has introduced a versatile functional form which subsumes a number of the more popular choices of divergence measures, including those mentioned above. For a convex function $f : [0,\infty) \to \mathbf{R}$, the *Csiszár f-divergence* between p and q is defined by

$$I_f(p,q) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right). \tag{1.5}$$

Thus the variational distance is given by $f(u) = |u-1|$, Kullback–Leibler divergence by $f(u) = u\ln u$, triangular discrimination by $f(u) = (u-1)^2/(u+1)$ and Hellinger discrimination by $f(u) = \frac{1}{2}(\sqrt{u}-1)^2$.

Most common choices of f, like the above, satisfy $f(1) = 0$, so that $I_f(p,p) = 0$. Convexity then ensures that $I_f(p,q)$ is nonnegative. However as noted in [11], some additional flexibility for applications can be achieved by not insisting on convexity.

Closely related to the above are questions of approximation. For example, in practice exact values for each p_i and q_i may not be available, and it is desirable to be able to give an estimate for $I_f(p,q)$ or to impose upper bounds on its value with only limited knowledge of p and q. To this end a recent paper [11] derived an approximation theorem by the use of Ostrowski's integral inequality for absolutely continuous functions. It is convenient to invoke as a benchmark the chi–squared discrepancy measure

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

which arises from (1.5) as the particular case $f(u) = (u-1)^2$. The main result derived in [11] is as follows.

Theorem A. *Suppose that there exist real numbers r, R with*

$$0 < r \le p_i/q_i \le R < \infty \text{ for all } i \in \{1,...,n\}. \tag{1.6}$$

Assume that $f : [r, R] \to \mathbf{R}$ *is absolutely continuous on* $[r, R]$ *and* $f' \in L_\infty[r, R]$, *that is, that*

$$\left\| f' \right\|_\infty := ess \sup_{t\in[r,R]} \left| f'(t) \right| < \infty.$$

Then

$$\begin{aligned}
&\left| I_f(p,q) - \frac{1}{R-r} \int_r^R f(t)dt \right| \\
&\quad \le \left[\frac{1}{4} + \frac{1}{(R-r)^2} \left\{ D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1 \right)^2 \right\} \right] (R-r) \left\| f' \right\|_\infty \\
&\quad \le \frac{1}{2}(R-r) \left\| f' \right\|_\infty .
\end{aligned}$$

Suppose that p and q are close in the sense that $\varepsilon := R - r$ is small. Then Theorem A provides an approximation for the Csiszár f–divergence of accuracy of order $O(\varepsilon)$ with only a mild assumption on f'. However, common choices for f in the engineering literature are frequently quite smooth, so that stronger assumptions are not necessarily restrictive. We can expect that, with appropriate restrictions, approximations to the Csiszár f–divergence can be obtained with higher–order accuracy in ε. In this article we derive approximations of order of accuracy $O(\varepsilon^2)$ and $O(\varepsilon^3)$.

We note at the outset that it would appear that a programme of obtaining such approximations can only to a limited extent be achieved in line with the goal in [11] of employing approximants to $I_f(p,q)$ that involve p, q through r and R alone. This can achieved with accuracy order $O(\varepsilon^2)$ but not in general with accuracy order $O(\varepsilon^3)$.

In Section 2 we discuss some preliminary ideas and state and prove a general proposition useful in questions concerning approximations. In Section 3 we establish the basic results and in Section 4 address as examples some of the more common choices of f. We conclude in Section 5 with an application to mutual information in information theory.

2. Preliminaries

As an introduction to the ideas, suppose f' is absolutely continuous on $[r, R]$ and $f'' \in L_\infty[r, R]$. Then for $x \in [r, R]$ we have

$$f(x) = f(1) + (x-1)f'(1) + \frac{(x-1)^2}{2} f''(\xi_x)$$

for some $\xi_x \in [r, R]$. Setting $x = p_i/q_i$ gives for each $i = 1, \ldots, n$ that

$$q_i f(p_i/q_i) = q_i f(1) + (p_i - q_i)f'(1) + \frac{(p_i - q_i)^2}{2q_i} f''(\xi_i)$$

for some $\xi_i \in [r, R]$. Since $\sum_{i=1}^n p_i = 1 = \sum_{i=1}^n q_i$, summation over i yields

$$I_f(p,q) = f(1) + \frac{1}{2} \sum_{i=1}^n \frac{(p_i - q_i)^2}{q_i} f''(\xi_i),$$

from which we deduce that

$$|I_f(p,q) - f(1)| \leq \frac{1}{2} D_{\chi^2}(p,q) \left\| f^{''} \right\|_\infty .$$

It was shown in [11] that

$$D_{\chi^2}(p,q) \leq (R-1)(1-r). \tag{2.1}$$

Also the choices $\alpha = R-1$, $\beta = 1-r$ in the elementary inequality

$$\alpha\beta \leq \frac{1}{4}(\alpha+\beta)^2$$

give $(R-1)(1-r) \leq (R-r)^2/4$, so that (2.1) can be extended to

$$D_{\chi^2}(p,q) \leq (R-1)(1-r) \leq (R-r)^2/4. \tag{2.2}$$

This leads to

$$|I_f(p,q) - f(1)| \leq \frac{\varepsilon^2}{8} \left\| f^{''} \right\|_\infty , \tag{2.3}$$

in which the approximant does not depend on p, q. This approximation is, however, of limited utility, since as noted $f(1) = 0$ for many common choices of f.

A similar argument applies when $f^{''}$ is absolutely continuous on $[r,R]$ and $f^{'''} \in L_\infty[r,R]$. Paralleling the line of reasoning above leads to

$$I_f(p,q) = f(1) + \frac{1}{2}\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} f^{''}(1) + \frac{1}{6}\sum_{i=1}^{n} \frac{(p_i - q_i)^3}{q_i^2} f^{'''}(\eta_i),$$

where each $\eta_i \in [r,R]$, whence

$$\left| I_f(p,q) - f(1) - \frac{1}{2} D_{\chi^2}(p,q) f^{''}(1) \right| \leq \frac{1}{6} D_{|\chi|^3}(p,q) \left\| f^{'''} \right\|_\infty .$$

Here

$$D_{|\chi|^3}(p,q) := \sum_{i=1}^{n} \frac{|p_i - q_i|^3}{q_i^2}$$

is the chi–cubed discrepancy, which corresponds to $f(u) = (u-1)^3$.

It follows at once from (1.6) and $\sum_{i=1}^{n} p_i = 1 = \sum_{i=1}^{n} q_i$ that $r \leq 1 \leq R$. Hence

$$\left| \frac{p_i}{q_i} - 1 \right| \leq \max(1-r, R-1) \leq R-r \text{ for } i = 1, \ldots, n, \tag{2.4}$$

and from (2.2) we have

$$D_{|\chi|^3}(p,q) \leq (R-r) D_{\chi^2}(p,q) \leq (R-r)(R-1)(1-r) \leq \frac{1}{4}(R-r)^3. \tag{2.5}$$

Therefore

$$\left| I_f(p,q) - f(1) - \frac{1}{2} D_{\chi^2}(p,q) f^{''}(1) \right| \leq \frac{\varepsilon^3}{24} \left\| f^{'''} \right\|_\infty . \tag{2.6}$$

This time the approximant involves p, q nontrivially.

In this article we establish results similar to (2.3) and (2.6) but involving approximants for $I_f(p,q)$ for which we can achieve tighter error bounds. Our basic tool is Ostrowski's integral inequality for absolutely continuous mappings, Theorem B below. For a short proof and applications to numerical integration, the reader is referred to [8].

Theorem B. *Assume that $g : [a,b] \to \mathbf{R}$ is absolutely continuous with $g^{'} \in L_\infty[a,b]$. Then*

$$\left| g(x) - \frac{1}{b-a} \int_a^b g(t)dt \right| \le \left[\frac{1}{4} + \left(\frac{x - \frac{a+b}{2}}{b-a} \right)^2 \right] (b-a) \left\| g^{'} \right\|_\infty$$

for all $x \in [a,b]$.

We shall also make use of a result which both generalizes and improves (2.5). This concerns the divergence measure

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

with $m \ge 1$.

Proposition 1. *Suppose that (1.6) is satisfied with $r < R$ and that $m \ge 1$. Then*

$$D_{|\chi|^m}(p,q) \le \frac{(R-1)(1-r)}{R-r} \left[(1-r)^{m-1} + (R-1)^{m-1} \right] \le \left(\frac{R-r}{2} \right)^m.$$

The first inequality is an equality if and only if p, q form a boundary pair with respect to r and R, that is, for each i either $p_i/q_i = r$ or $p_i/q_i = R$. The second inequality is an equality if and only if $R + r = 2$, that is, r and R are equidistant from unity.

Proof. Suppose that there exists a value of $i \in \{1, \dots, n\}$ for which $p_i/q_i = \theta$ with $r < \theta < R$. We may put

$$p_i = r\frac{R-\theta}{R-r} q_i + R\frac{\theta - r}{R-r} q_i = p_{i,1} + p_{i,2},$$

say, and similarly

$$q_i = \frac{R-\theta}{R-r} q_i + \frac{\theta - r}{R-r} q_i = q_{i,1} + q_{i,2}.$$

We then have

$$p_{i,1}/q_{i,1} = r \text{ and } p_{i,2}/q_{i,2} = R.$$

We may replace the probability n–vectors p, q with probability $(n+1)$–vectors by replacing p_i by the ordered pair $p_{i,1}, p_{i,2}$ and q_i by the ordered pair $q_{i,1}, q_{i,2}$.

The contribution to $D_{|\chi|^m}(p,q)$ from p_i and q_i is

$$\frac{|p_i - q_i|^m}{q_i^{m-1}} = q_i |1-\theta|^m.$$

The contribution from $p_{i,1}$, $q_{i,1}$ and $p_{i,2}$, $q_{i,2}$ after such a replacement is

$$\frac{R-\theta}{R-r} q_i (1-r)^m + \frac{\theta - r}{R-r} q_i (R-1)^m.$$

We now show that the change gives rise to an increase in the divergence $D_{|\chi|^m}(p,q)$. We need to prove that

$$\phi(\theta) := \frac{R-\theta}{R-r}(1-r)^m + \frac{\theta - r}{R-r}(R-1)^m - |1-\theta|^m$$

is positive if $r < \theta < R$.

First suppose that $m > 1$. Since

$$\phi^{''}(\theta) = -m(m-1)(1-\theta)^{m-2} \text{ for } r < \theta < 1,$$

ϕ is strictly concave on $(r, 1)$. Also $\phi(r) = 0$ and $\phi(1) > 0$, so that $\phi(\theta)$ is positive on $(r, 1]$. Similarly

$$\phi^{''}(\theta) = -m(m-1)(\theta-1)^{m-2} \text{ for } 1 < \theta < R,$$

so ϕ is strictly concave on $(1, R)$. Since $\phi(1) > 0$ and $\phi(R) = 0$, ϕ is positive on $[1, R)$. Thus $\phi(\theta) > 0$ on (r, R).

Now suppose $m = 1$. We have

$$\phi^{'}(\theta) = 2\frac{R-1}{R-r} > 0 \text{ for } r < \theta < 1$$

with $\phi(r) = 0$, so ϕ is positive on $(r, 1]$. Further

$$\phi^{'}(\theta) = -2\frac{1-r}{R-r} < 0 \text{ for } 1 < \theta < R$$

with $\phi(R) = 0$, so ϕ is also positive on $[1, R)$. Thus we have again that $\phi(\theta) > 0$ on (r, R).

In either case we have demonstrated the claimed increase in the divergence. It follows that if we seek an upper bound to $D_{|\chi|^m}(p, q)$, we can without loss of generality restrict attention to boundary pairs p, q. After relabelling if necessary, we assume that

$$p_i/q_i = r \text{ for } i = 1, \ldots, k$$

and

$$p_i/q_i = R \text{ for } i = k+1, \ldots, n.$$

Then

$$1 = \sum_{i=1}^{n} p_i = r\sum_{i=1}^{k} q_i + R\sum_{i=k+1}^{n} q_i = r\sum_{i=1}^{k} q_i + R\left[1 - \sum_{i=1}^{k} q_i\right]$$

and we have

$$\sum_{i=1}^{k} q_i = \frac{R-1}{R-r}, \qquad \sum_{i=k+1}^{n} q_i = \frac{1-r}{R-r}.$$

Thus for a boundary pair p, q we have

$$\begin{aligned} D_{|\chi|^m}(p, q) &= \sum_{i=1}^{k} \frac{(q_i - p_i)^m}{q_i^{m-1}} + \sum_{i=k+1}^{n} \frac{(p_i - q_i)^m}{q_i^{m-1}} \\ &= (1-r)^m \sum_{i=1}^{k} q_i + (R-1)^m \sum_{i=k+1}^{n} q_i \\ &= \frac{R-1}{R-r}(1-r)^m + \frac{1-r}{R-r}(R-1)^m. \end{aligned}$$

Now by elementary calculus we have that for $x \geq 0$, $y \geq 0$ with $x + y = 2c$, a constant, the function $yx^m + xy^m$ takes its maximum value $2c^{m+1}$ uniquely when $x = y = c$. Applying this with $x = 1 - r$, $y = R - 1$ gives that

$$\frac{R-1}{R-r}(1-r)^m + \frac{1-r}{R-r}(R-1)^m \leq \left(\frac{R-1}{2}\right)^m,$$

with equality occurring when $R - 1 = 1 - r$. □

The main part of the result for the case $m = 2$ is simply (2.2). The result for $m = 3$ enables us to strengthen (2.6) to

$$\left| I_f(p,q) - f(1) - \frac{1}{2} D_{\chi^2}(p,q) f^{''}(1) \right| \leq \frac{\varepsilon^3}{48} \left\| f^{'''} \right\|_\infty .$$

We shall, however, look for a much closer approximant for $I_f(p,q)$.

3. Basic results

We assume in what follows that there exist real numbers r, R satisfying (1.6). As noted in the introduction this entails that $r \leq 1 \leq R$.

For f satisfying the conditions of Theorem A, we define $f^* : [r,R] \to \mathbf{R}$ by

$$f^*(u) := f(1) + (u-1) f^{'} \left(\frac{1+u}{2} \right).$$

We remark that this gives $f^*(1) = f(1)$.

Theorem 1. *Assume that $f : [0,\infty) \to \mathbf{R}$ is such that $f^{'} : [r,R] \to \mathbf{R}$ is absolutely continuous on $[r,R]$ and $f^{''} \in L_\infty[r,R]$. Then*

$$\begin{aligned} |I_f(p,q) - I_{f^*}(p,q)| &\leq \frac{1}{4} \left\| f^{''} \right\|_\infty D_{\chi^2}(p,q) \\ &\leq \frac{1}{4} \left\| f^{''} \right\|_\infty (R-1)(1-r) \\ &\leq \frac{1}{16} \left\| f^{''} \right\|_\infty (R-r)^2. \end{aligned} \tag{3.1}$$

Proof. Taking $x = (a+b)/2$ in Theorem B yields the midpoint inequality

$$\left| g\left(\frac{a+b}{2} \right)(b-a) - \int_a^b g(t)dt \right| \leq \frac{1}{4}(b-a)^2 \left\| g^{'} \right\|_\infty .$$

The choices $g = f^{'}$, $a = 1$ and $b = x \in [r,R]$ provide

$$\left| f(x) - f(1) - (x-1) f^{'} \left(\frac{1+x}{2} \right) \right| \leq \frac{1}{4}(x-1)^2 \left\| f^{''} \right\|_\infty$$

for all $x \in [r,R]$.

Putting $x = p_i/q_i \in [r,R]$ gives

$$\left| q_i f\left(\frac{p_i}{q_i} \right) - q_i f(1) - (p_i - q_i) f^{'} \left(\frac{p_i + q_i}{2q_i} \right) \right| \leq \frac{1}{4} \left(\frac{p_i - q_i}{q_i} \right)^2 q_i \left\| f^{''} \right\|_\infty$$

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

Summing over i and using the generalized triangle inequality leads to

$$\left| I_f(p,q) - f(1) - \sum_{i=1}^n (p_i - q_i) f^{'} \left(\frac{p_i + q_i}{2q_i} \right) \right| \leq \frac{1}{4} \left\| f^{''} \right\|_\infty \sum_{i=1}^n \frac{(p_i - q_i)^2}{q_i},$$

whence we have the first inequality in (3.1).

The second and third inequalities follow from (2.2). □

Corollary 1. *Let f be as in Theorem 1. If $\varepsilon > 0$ and*

$$0 \le R - r \le 4 \cdot \sqrt{\varepsilon / \|f''\|_\infty},$$

then

$$|I_f(p,q) - I_{f^*}(p,q)| \le \varepsilon.$$

A second corollary emphasizes the approximation aspect for distribution p and q which are close.

Corollary 2. *Let $f : [0,2] \to \mathbf{R}$ be such that the derivative $f^{'} : [0,2] \to \mathbf{R}$ is absolutely continuous and $f^{''} \in L_\infty[0,2]$. If $\eta \in (0,1)$ and $p(\eta)$, $q(\eta)$ are such that*

$$\left| \frac{p_i(\eta)}{q_i(\eta)} - 1 \right| \le \eta \quad \textit{for all } i \in \{1, ..., n\}, \tag{3.2}$$

then

$$I_f\,(p(\eta), q(\eta)) = I_{f^*}\,(p(\eta), q(\eta)) + R_f(p,q,\eta) \tag{3.3}$$

and the remainder $R_f(p,q,\eta)$ satisfies the estimate

$$R_f(p,q,\eta) \le \frac{1}{4} \left\|f^{''}\right\|_\infty D_{\chi^2}(p(\eta), q(\eta)) \le \frac{1}{4} \left\|f^{''}\right\|_\infty \eta^2. \tag{3.4}$$

Proof. Set $R = 1 + \eta$ and $r = 1 - \eta$ in Theorem 1. □

Theorem 2. *If $f : [a,b] \to \mathbf{R}$ is such that $f^{''}$ is absolutely continuous and $f^{'''} \in L_\infty[r,R]$, then*

$$\begin{aligned} |I_f\,(p,q) - I_{f^*}\,(p,q)| &\le \frac{1}{24} \left\|f^{'''}\right\|_\infty D_{|\chi|^3}(p,q) \\ &\le \frac{1}{24} \left\|f^{'''}\right\|_\infty \frac{(R-1)(1-r)}{R-r} \left[(1-r)^2 + (R-1)^2\right] \\ &\le \frac{1}{192} \left\|f^{'''}\right\|_\infty (R-r)^3. \end{aligned} \tag{3.5}$$

The constants are best–possible.

Proof. For $g : [a,b] \to \mathbf{R}$ such that $g^{'}$ is absolutely continuous on $[a,b]$ and $g^{''} \in L_\infty[r,R]$, we have the midpoint inequality

$$\left| g\left(\frac{a+b}{2}\right)(b-a) - \int_a^b g\,(t)\,dt \right| = \frac{1}{24}\,(b-a)^3 \left\|g^{''}\right\|_\infty$$

arising in numerical integration. Setting $g = f^{'}$, $a = 1$ and $b = x \in [r,R]$ supplies

$$\left| f(x) - f(1) - (x-1) f^{'}\left(\frac{1+x}{2}\right) \right| = \frac{1}{24}\,|x-1|^3 \left\|f^{'''}\right\|_\infty$$

for all $x \in [r,R]$. We may proceed as in the proof of Theorem 1 to deduce the first inequality in (3.5). Proposition 1 for $m = 3$ gives the others. To complete the proof it suffices to prove that the constant $1/24$ in the first inequality is best–possible.

Choose $f(u) = (u-1)^3$, so that $I_f(p,q) = D_{|\chi|^3}(p,q)$. Then

$$f^*(u) = \frac{3}{4}(u-1)^3 = \frac{3}{4} f(u),$$

so that

$$|I_f(p,q) - I_{f^*}(p,q)| = \frac{1}{4} D_{|\chi|^3}(p,q).$$

As $\left\| f^{'''} \right\|_\infty = 6$, the first inequality in (3.5) for this choice of f is thus an equality.

□

Corollary 3. *Let f be as in Theorem 2. If $\varepsilon > 0$ and*

$$0 \le R - r \le 4 \cdot \sqrt[3]{3\varepsilon / \|f'''\|_\infty},$$

then

$$|I_f(p,q) - I_{f^*}(p,q)| \le \varepsilon.$$

Also, the following approximation result holds.

Corollary 4. *Let $f : [0,2] \to \mathbf{R}$ be so that $f^{'''} \in L_\infty[0,2]$. If $\eta \in (0,1)$ and $p(\eta)$, $q(\eta)$ satisfy (3.2), then we have the representation (3.3) and the remainder $R_f(p,q,\eta)$ satisfies the estimate*

$$|R_f(p,q,\eta)| \le \frac{1}{24} \left\| f^{'''} \right\|_\infty \eta^3.$$

Proof. The result follows from Theorem 2 with $r = 1 - \eta$ and $R = 1 + \eta$. □

4. Applications to some common divergence measures

When $f : (0,\infty) \to \mathbf{R}$ is the convex map $f(u) = u \ln u$, $I_f(p,q)$ becomes the Kullback–Leibler distance $D(p,q)$. We denote $I_{f^*}(p,q)$ by $D^*(p,q)$ and adopt a similar notation for other specific divergences. We have

$$\begin{aligned} D^*(p,q) &= \sum_{i=1}^n (p_i - q_i) \left[\ln\left(\frac{p_i + q_i}{2q_i} \right) + 1 \right] \\ &= \sum_{i=1}^n (p_i - q_i) \ln\left(\frac{p_i + q_i}{2q_i} \right). \end{aligned}$$

As $f^{''}(u) = 1/u$, we have

$$\left\| f^{''} \right\|_\infty = \sup_{u \in [r,R]} \left| f^{''}(u) \right| = 1/r,$$

and the conclusion of Theorem 1 becomes

$$|D(p,q) - D^*(p,q)| \le \frac{1}{4r} D_{\chi^2}(p,q) \le \frac{(R-1)(1-r)}{4r} \le \frac{(R-r)^2}{16r}. \tag{4.1}$$

An example of a concave $f : (0,\infty) \to \mathbf{R}$ with $f(1) = 0$ is $f(u) = \ln u$. We have

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

and

$$-D^*(q,p) = \sum_{i=1}^n (p_i - q_i) \frac{2q_i}{p_i + q_i} = 2 \sum_{i=1}^n q_i \cdot \frac{p_i - q_i}{p_i + q_i}.$$

As $f^{''}(u) = -1/u^2$, we have $\left\| f^{''} \right\|_\infty = 1/r^2$.

Consequently (3.1) reads

$$|D(q,p) - D^*(q,p)| \le \frac{1}{4r^2} D_{\chi^2}(p,q) \le \frac{(R-1)(1-r)}{4r^2} \le \frac{1}{16}\left(\frac{R}{r}-1\right)^2. \tag{4.2}$$

For $f : (0,\infty) \to \mathbf{R}$ given by $f(u) = \frac{1}{2}\left(\sqrt{u}-1\right)^2$, the f–divergence $I_f(p,q)$ becomes the Hellinger discrimination and

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

As $f^{''}(u) = 1/(4u^{2/3})$ for $u \in (0,\infty)$,

$$\left\|f^{''}\right\|_\infty = \sup_{u\in[r,R]}\left|f^{''}(u)\right| = \frac{1}{4r^{2/3}}.$$

By Theorem 1 we have

$$\left|h^2(p,q) - h^{2*}(p,q)\right| \le \frac{1}{16r^{2/3}} D_{\chi^2}(p,q) \le \frac{(R-1)(1-r)}{16r^{2/3}} \le \frac{1}{64r^{2/3}}(R-r)^2.$$

The Renyi α–order distance $R_\alpha(p,q) := \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha}$ is given by $f : (0,\infty) \to \mathbf{R}$ with $f(u) = u^\alpha$ $(\alpha > 1)$. This provides an example in which f may or may not be convex and $f(1) \ne 0$. We have

$$R_\alpha^*(p,q) = 1 + \frac{\alpha}{2^{\alpha-1}}\sum_{i=1}^{n}(p_i - q_i)\left(\frac{p_i+q_i}{q_i}\right)^{\alpha-1},$$

and as $f^{''}(u) = \alpha(\alpha-1)u^{\alpha-2}$, we have

$$\left\|f^{''}\right\|_\infty = \delta_\alpha(r,R) := \begin{cases} \alpha(\alpha-1)R^{\alpha-2} & \text{if } 2 \le \alpha < \infty \\ \alpha(\alpha-1)r^{\alpha-2} & \text{if } 1 < \alpha \le 2. \end{cases}$$

Theorem 1 states that

$$\begin{aligned} |R_\alpha(p,q) - R_\alpha^*(p,q)| &\le \frac{1}{4}\delta_\alpha(r,R) D_{\chi^2}(p,q) \\ &\le \frac{1}{4} \times \begin{cases} \alpha(\alpha-1)R^{\alpha-2}(R-1)(1-r) & \text{if } 2 \le \alpha < \infty \\ \alpha(\alpha-1)r^{\alpha-2}(R-1)(1-r) & \text{if } 1 < \alpha \le 2. \end{cases} \end{aligned}$$

We now turn to some bounds provided by Theorem 2.

Reconsider the mapping $f : (0,\infty) \to \mathbf{R}$ with $f(u) = u \ln u$. We have $f^{'''}(u) = -1/u^2$, $\left\|f^{'''}\right\|_\infty = 1/r^2$ and by Theorem 2

$$|D(p,q) - D^*(p,q)| \le \frac{1}{24r^2} D_{|\chi|^3}(p,q) \le \frac{1}{192r^2}(R-r)^3.$$

For $f(u) = \ln u$, we have $f^{'''}(u) = 2/(3u^3)$ for $u \in [r,R]$ and then $\left\|f^{'''}\right\|_\infty = 2/(3r^3)$. By Theorem 2 we have

$$|D(q,p) - D^*(q,p)| \le \frac{1}{36r^3} D_{|\chi|^3}(p,q) \le \frac{1}{288}\left(\frac{R}{r}-1\right)^3.$$

For $f:(0,\infty)\to\mathbf{R}$ with $f(u)=\frac{1}{2}\left(\sqrt{u}-1\right)^2$, we have $f^{'''}(u)=-\frac{1}{6}\cdot u^{-5/3}$ for $u\in[r,R]$ and so $\left\|f^{'''}\right\|_\infty=\frac{1}{6}\cdot r^{-5/3}$. Consequently Theorem 2 gives

$$\left|h^2(q,p)-h^{2*}(q,p)\right|\leq\frac{1}{144r^{5/3}}D_{|\chi|^3}(p,q)\leq\frac{1}{1152r^{5/3}}\left(R-r\right)^3.$$

Finally for $f:(0,\infty)\to\mathbf{R}$ with $f(u)=u^\alpha$ $(\alpha>1)$, we have $f^{'''}(u)=\alpha(\alpha-1)(\alpha-2)u^{\alpha-3}$ and

$$\left\|f^{'''}\right\|_\infty=\eta_\alpha(r,R)=\begin{cases}\alpha(\alpha-1)(\alpha-2)R^{\alpha-3} & \text{if } 3\leq\alpha<\infty\\ \alpha(\alpha-1)(\alpha-2)r^{\alpha-3} & \text{if } 1<\alpha\leq 3.\end{cases}$$

Consequently Theorem 2 provides

$$\begin{aligned}|R_\alpha(p,q)-R_\alpha^*(p,q)| &\leq\frac{1}{24}\eta_\alpha(r,R)D_{|\chi|^3}(p,q)\\ &\leq\frac{1}{192}\times\begin{cases}\alpha(\alpha-1)(\alpha-2)R^{\alpha-2}(R-r)^3 & \text{if } 3\leq\alpha<\infty\\ \alpha(\alpha-1)(\alpha-2)r^{\alpha-2}(R-r)^3 & \text{if } 1<\alpha\leq 3.\end{cases}\end{aligned}$$

5. Application to mutual information

We consider mutual information, which is a measure of the amount of information one random variable provides about another. It is the reduction of uncertainty about one variable due to the knowledge of the other.

Definition 1. *Consider two random variables X and Y with a joint probability mass function $t(x,y)$ and marginal probability mass functions $p(x)$ $(x\in\mathcal{X})$ and $q(y)$ $(y\in\mathcal{Y})$. The mutual information is the relative entropy between the joint distribution and the product distribution, that is,*

$$I(X;Y)=\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}}t(x,y)\ln\left[\frac{t(x,y)}{p(x)q(y)}\right]=D(t(x,y),p(x)q(y))$$

where $D(\cdot,\cdot)$ denotes Kullback–Leibler distance.

We assume in what follows that

$$s\leq\frac{t(x,y)}{p(x)q(y)}\leq S\quad\text{for all } (x,y)\in\mathcal{X}\times\mathcal{Y}. \tag{5.1}$$

We have immediately that $s\leq 1\leq S$.

We may also define mutual information in a chi–squared sense, that is,

$$I_{\chi^2}(X;Y)=\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}}\frac{t^2(x,y)}{p(x)q(y)}-1.$$

Using inequality (4.1), we have the following proposition.

Proposition 2. *If t, p, q satisfy (5.1), we have*

$$\begin{aligned}&\left|I(X;Y)-\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}}[t(x,y)-p(x)q(y)]\ln\left[\frac{t(x,y)+p(x)q(y)}{2p(x)q(y)}\right]\right|\\ &\leq\frac{1}{4s}I_{\chi^2}(X;Y)\leq\frac{1}{4s}(S-1)(1-s)\leq\frac{1}{16s}(S-s)^2.\end{aligned}$$

A similar result follows from (4.2).

References

[1] F. Topsoe, Some inequalities for information divergence and related measures of discrimination, *RGMIA Res. Rep. Coll.,* **2** (1999) No. 1, 85–98.
[2] I. Csiszár, A note on Jensen's inequality, *Studia Sci. Math. Hung.,* **1** (1966) 185–188.
[3] I. Csiszár, Information-type measures of differences of probability distributions and indirect observation, *Studia Sci. Math. Hung.,* **2** (1967) 299–318.
[4] I. Csiszár, On topological properties of f-divergence, *Studia Sci. Math. Hung.,* **2** (1967) 329–339.
[5] L. LeCam, *Asymptotic Methods in Statistical Decision Theory*, New York: Springer, 1986.
[6] D. Dacunha-Castelle, *Ecole d'été de probabilités de Saint–Fleur*, VII–1977, Berlin, Heidelberg New York: Springer, 1978.
[7] C. Kraft, Some conditions of consistency and uniform consistency of statistical procedure, *Univ. of California Publ. in Statistics, Ph.D. Thesis*, **1** (1955), 125–142.
[8] S. S. Dragomir and S. Wang, Application of Ostrowski's inequality to the estimation of error bounds for some special means and for some numerical quadrature rules, *Appl. Math. Lett.,* ***11*** *(1998),* 105-109.
[9] D. S. Mitrinović, *Analytic Inequalities*, Springer Verlag, 1970.
[10] T. M. Cover and J. A. Thomas*, Elements of Information Theory*, John Wiley & Sons Inc., 1991.
[11] S. S. Dragomir, V. Gluščević and C. E. M. Pearce, Csiszár's f–divergence, Ostrowski's inequality and mutual information, *submitted*.

School of Communications and Informatics, Victoria University, PO Box 14428, MCMC, Melbourne, Victoria 8001, Australia
E-mail address: `sever@matilda.vu.edu.au`

Royal Australian Air Force, ARDU, PO Box 1500, Salisbury SA 5108, Australia
E-mail address: `vgluscev@maths.adelaide.edu.au`

Applied Mathematics Department, The University of Adelaide, Adelaide, SA 5005, Australia
E-mail address: `cpearce@maths.adelaide.edu.au`