CSISZÁR f–DIVERGENCE, OSTROWSKI'S INEQUALITY AND MUTUAL INFORMATION

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

Abstract. The Ostrowski integral inequality for an absolutely continuous function is used to provide a simple approximation to Csiszár's f–divergence measure for the difference between two probability distributions defined on a finite set. Concrete examples are given and some applications made to mutual information.

1. Introduction

The difference between two probability measures p, q on a set $A = \{\alpha_i | 1 \le i \le n\}$ is commonly measured in a variety of ways. Denote by p_i, q_i the associated point probabilities for the event $\alpha_i \in A$. To avoid triviality we assume that $p_i + q_i > 0$ for each i. The *variational distance* (ℓ_1–distance) and *information divergence* (Kullback–Leibler divergence) between the distributions p and q are defined respectively by

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

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

Another measure, which proves a useful benchmark in our analysis, is the chi–squared divergence of p, q, which is defined by

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

The last two measures are unfortunately infinite if $p_i > 0$ but $q_i = 0$ for some i. This complication is obviated in the *triangular discrimination* between p and q, which is defined as in [10] by

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

A generalization of this measure, parameterized by a natural number υ, is

$$\Delta_\upsilon(p,q) := \sum_{i=1}^{n} \frac{|p_i - q_i|^{2\upsilon}}{(p_i + q_i)^{2\upsilon - 1}},$$

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

Key words and phrases. Csiszár f–divergence, Ostrowski's inequality, divergence measure, mutual information.

which we refer to as *triangular discrimination of order* υ (see [10]). Another common choice is the *Hellinger discrimination*

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

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

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

and

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

(see LeCam [8] and Dacunha–Castelle [5]). 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.* [10]).

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 [5]). It follows from this that

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

Again the coefficient 2 is best–possible (see [10]).

The key to unity in this diversity is that all the discrepancy measures considered above are particular instances of Csiszár f–divergences. If $f:[0,\infty)\to\mathbf{R}$ is convex, the *Csiszár* f*-divergence* between p and q is defined by

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

(see Csiszár [2]–[4]). Thus the family $(f_s)_{s\ge 1}$ of functions with

$$f_s(u) = |u-1|^s(u+1)^{1-s}$$

gives rise to variational distance when $s=1$, triangular discrimination when $s=2$ and triangular discrimination of order υ when $s=2\upsilon$ (see [10]). The choice $f(u) = \frac{1}{2}(\sqrt{u}-1)^2$ gives rise to Hellinger discrimination and $f(u)=u\ln u$ to Kullback-Leibler divergence. The chi–squared divergence is given by $f(u)=(u-1)^2$.

For all of the above choices $f(1)=0$, so that $I_f(p,p)=0$. The convexity of f then ensures that $I_f(p,q)$ is nonnegative.

In Section 2 we derive, by the use of Ostrowski's integral inequality for absolutely continuous mappings with essentially bounded first derivative, an approximation for the Csiszár f–divergence in terms of an integral mean. With many concrete examples this provides very simple approximations. Section 3 considers some of the examples noted above and Section 4 the case when each pair p_i, q_i are very close. Finally, in Section 5, we look at applications to mutual information.

It needs to be stressed that as these estimates lose most of the detailed information involved in the values p_i, q_i, the approximations, while very simple, can also be very crude.

2. An inequality for Csiszár f–divergence

In some applications it is convenient to make use of definition (1.1) for functions $f : [0, \infty) \to \mathbf{R}$ which are continuous but not necessarily convex. An illustrative example is given in Section 3. Accordingly our main result, Theorem 1 below, does not assume convexity.

We assume in what follows that there exist real numbers r, R with

$$0 < r \leq p_i/q_i \leq R < \infty$$

for all $i \in \{1, ..., n\}$. Note that if $r > 1$, then $p_i > q_i$ for each i, which gives $1 = \sum_i p_i > \sum_i q_i = 1$, a contradiction. Hence $r \leq 1$. A similar argument gives $R \geq 1$.

Further suppose that the restriction of f to the compact interval $[r, R]$ is absolutely continuous. We derive an approximation to the Csiszár f–divergence in terms of the integral mean of f over $[r, R]$. We shall show in Theorem 1 below that if p and q are close in the sense that $R - r$ is small, then the integral mean

$$\frac{1}{R-r} \int_r^R f(t) dt$$

approximates the Csiszár f–divergence to first order.

We make use of Ostrowski's integral inequality, which states the following. See [6] for a short proof and some applications to numerical integration and special means.

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

$$\left\| g' \right\|_\infty := ess \sup_{t \in [a,b]} \left| g'(t) \right| < \infty.$$

Then

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

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

A further key result is due to Diaz and Metcalf (see [9, p. 61]).

Theorem B. *Suppose $a_k (\neq 0)$ and b_k $(k = 1, ..., n)$ are real numbers satisfying $m \leq b_k/a_k \leq M$. Then*

$$\sum_{k=1}^n b_k^2 + mM \sum_{k=1}^n a_k^2 \leq (M + m) \sum_{k=1}^n a_k b_k.$$

Equality holds if and only if for each k either $b_k = ma_k$ or $b_k = Ma_k$.

We shall make use of a slight extension of this.

Proposition 1. *Suppose the conditions of the Diaz–Metcalf result hold and* $t_k > 0$ *for* $k = 1, ..., n$. *Then*

$$\sum_{k=1}^{n} t_k b_k^2 + mM \sum_{k=1}^{n} t_k a_k^2 \leq (M+m) \sum_{k=1}^{n} t_k a_k b_k.$$

Equality holds if and only if for each k *either* $b_k = ma_k$ *or* $b_k = Ma_k$.

Proof. We have for $k = 1, 2, \ldots, n$ that

$$(b_k/a_k - m)(M - b_k/a_k) t_k a_k^2 \geq 0.$$

The desired result follows on summation over k. □

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

$$\left| I_f(p,q) - \frac{1}{R-r} \int_r^R f(t)dt \right| \leq \left[\frac{1}{4} + \frac{1}{(R-r)^2} \left\{ D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1 \right)^2 \right\} \right] (R-r) \left\| f^{'} \right\|_\infty \leq \frac{1}{2}(R-r) \left\| f^{'} \right\|_\infty . \tag{2.1}$$

Proof. By Ostrowski's integral inequality, we have

$$\left| f\left(\frac{p_i}{q_i}\right) - \frac{1}{R-r} \int_r^R f(t)dt \right| \leq \left[\frac{1}{4} + \left(\frac{\frac{p_i}{q_i} - \frac{R+r}{2}}{R-r} \right)^2 \right] (R-r) \left\| f^{'} \right\|_\infty$$

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

We may multiply by q_i, sum the resultant inequalities and use the generalized triangle inequality to obtain

$$\begin{aligned} &\left| I_f(p,q) - \frac{1}{R-r} \int_r^R f(t)dt \right| \\ &\leq \sum_{i=1}^{n} q_i \left| f\left(\frac{p_i}{q_i}\right) - \frac{1}{R-r} \int_r^R f(t)dt \right| \\ &\leq \left[\frac{1}{4} + \frac{1}{(R-r)^2} \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - \frac{R+r}{2} \right)^2 \right] (R-r) \left\| f^{'} \right\|_\infty . \end{aligned}$$

Since

$$\begin{aligned} \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - \frac{R+r}{2} \right)^2 &= \sum_{i=1}^{n} \frac{p_i^2}{q_i} - (R+r) + \left(\frac{R+r}{2} \right)^2 \\ &\leq \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1 + \left(\frac{R+r}{2} - 1 \right)^2 \\ &= D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1 \right)^2 , \end{aligned}$$

this yields the first inequality in (2.1).

For the second, set $b_k = \sqrt{p_k/q_k}$ and $a_k = \sqrt{q_k/p_k}$ $(k = 1, ..., n)$. Then $a_k/b_k = p_k/q_k \in [r, R]$ $(k \in \{1, ..., n\})$. On applying Proposition 1 for $t_k = p_k$ $(k = 1, ..., n)$, we get

$$\sum_{k=1}^{n} p_k \left(\sqrt{\frac{p_k}{q_k}}\right)^2 + rR \sum_{k=1}^{n} p_k \left(\sqrt{\frac{q_k}{p_k}}\right)^2 \leq (r + R) \sum_{k=1}^{n} p_k \sqrt{\frac{p_k}{q_k}} \cdot \sqrt{\frac{q_k}{p_k}},$$

or equivalently

$$\sum_{k=1}^{n} \frac{p_k^2}{q_k} + rR \leq R + r.$$

Thus

$$D_{\chi^2}(p, q) \leq r + R - rR - 1 = (1 - r)(R - 1)$$

and so

$$\frac{1}{4} + \frac{1}{(R - r)^2} \left[D_{\chi^2}(p, q) + \frac{1}{4}(R + r - 2)^2\right] \leq \frac{1}{2}$$

and the theorem is proved. □

Corollary 1. *Let f satisfy the conditions of Theorem 1. If $\varepsilon > 0$ and*

$$0 \leq R - r \leq 2\epsilon / \left\| f^{'} \right\|_{\infty},$$

then

$$\left| I_f(p, q) - \frac{1}{R - r} \int_{r}^{R} f(t)dt \right| \leq \epsilon.$$

Theorem 1 can be reformulated to emphasize the approximation aspect.

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

$$\left| \frac{p_i(\eta)}{q_i(\eta)} - 1 \right| \leq \eta$$

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

$$I_f(p(\eta), q(\eta)) = \frac{1}{2\eta} \int_{1-\eta}^{1+\eta} f(t)dt + R_f(p, q, \eta)$$

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

$$|R_f(p, q, \eta)| \leq \frac{\eta}{2} \left[1 + \frac{1}{\eta^2} D_{\chi^2}(p(\eta), q(\eta))\right] \left\| f^{'} \right\|_{\infty} \leq \eta \left\| f^{'} \right\|_{\infty}.$$

This follows by Theorem 1 with the choices $R = 1 + \eta$ and $r = 1 - \eta$ $(\eta \in (0, 1))$.

3. Particular cases

For Kullback–Leibler distance, we take $f(u) = u \ln u$. With this choice we have $\left\| f' \right\|_\infty = \ln(eR)$ and

$$\int_r^R f(t)dt = \frac{1}{4}\left[R^2 \ln R^2 - r^2 \ln r^2 - \left(R^2 - r^2\right)\right]$$
$$= \frac{R^2 - r^2}{4} \ln \left[\left(\frac{(R^2)^{(R^2)}}{(r^2)^{(r^2)}} \right)^{1/(R^2 - r^2)} \cdot \frac{1}{e} \right]$$
$$= \frac{R^2 - r^2}{4} I\left[\left(R^2, r^2\right)\right],$$

where the identric mean $I(a, b)$ for positive arguments is given by

$$I(a,b) := \begin{cases} a & \text{if } b = a \\ \frac{1}{e}\left(\frac{b^b}{a^a}\right)^{1/(b-a)} & \text{if } b \neq a. \end{cases}$$

The conclusion of Theorem 1 reads

$$\left| D(p,q) - \frac{R+r}{4} \ln\left[I\left(R^2, r^2\right)\right] \right| \tag{3.1}$$
$$\leq \left[\frac{1}{4} + \frac{1}{(R-r)^2}\left[D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1\right)^2\right]\right](R-r)\ln(eR)$$
$$\leq \frac{1}{2}(R-r)\ln(eR).$$

If we take the concave map $f : (0, \infty) \to \mathbf{R}$ given by $f(u) = \ln u$, then we have

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

With this choice $\left\| f' \right\|_\infty = 1/r$ and the identric mean reappears through

$$\frac{1}{R-r}\int_r^R f(t)dt = \ln\left[I(r,R)\right].$$

Theorem 1 provides

$$\left| D(q,p) - \ln\left[\frac{1}{I(r,R)}\right] \right| \tag{3.2}$$
$$\leq \left[\frac{1}{4} + \frac{1}{(R-r)^2}\left[D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1\right)^2\right]\right]\left(\frac{R}{r} - 1\right)$$
$$\leq \frac{1}{2}\left(\frac{R}{r} - 1\right).$$

For Hellinger discrimination $f(u) = \left(\sqrt{u} - 1\right)^2/2$, so

$$f'(u) = \frac{\sqrt{u} - 1}{2\sqrt{u}}, \quad f''(u) = \frac{1}{4u}$$

for $u \in (0, \infty)$ and

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

Also

$$\frac{1}{R-r} \int_{r}^{R} f(t)dt = \frac{R+r}{4} - \frac{2}{3} \cdot \frac{R + \sqrt{rR} + r}{\sqrt{r} + \sqrt{R}} + \frac{1}{2},$$

and inequality (2.1) becomes

$$\begin{aligned} &\left| h^2(p,q) - \left[\frac{R+r}{4} - \frac{2}{3} \cdot \frac{R + \sqrt{rR} + r}{\sqrt{r} + \sqrt{R}} + \frac{1}{2} \right] \right| \\ &\quad \leq \left[\frac{1}{4} + \frac{1}{(R-r)^2} \left\{ D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1 \right)^2 \right\} \right] (R-r) \left[\frac{\sqrt{R} - 1}{2\sqrt{R}} \right] \\ &\quad \leq \frac{1}{4\sqrt{R}} (R-r)(\sqrt{R} - 1). \end{aligned}$$

For variational distance, $f(u) = |u - 1|$, which is absolutely continuous on $[r, R]$. We have

$$f^{'}(u) := \begin{cases} -1 & \text{if } u \in (r,1) \\ 1 & \text{if } u \in (1,R), \end{cases}$$

so that

$$\left\| f^{'} \right\|_{\infty} = \sup_{t \in [r,R]} |f^{'}(t)| = 1.$$

Further

$$\begin{aligned} \frac{1}{R-r} \int_{r}^{R} f(t)dt &= \frac{1}{R-r} \left[\int_{r}^{1} (1-u)du + \int_{1}^{R} (u-1)du \right] \\ &= \frac{1}{R-r} \left[\frac{(r-1)^2}{2} + \frac{(R-1)^2}{2} \right] \\ &= \frac{1}{R-r} \left[\frac{(R-r)^2}{4} + \left(\frac{r+R}{2} - 1 \right)^2 \right]. \end{aligned}$$

Theorem 1 provides

$$\begin{aligned} &\left| V(p,q) - \frac{1}{R-r} \left[\frac{(R-r)^2}{4} + \left(\frac{r+R}{2} - 1 \right)^2 \right] \right| \\ &\quad \leq \left[\frac{1}{4} + \frac{1}{(R-r)^2} \left\{ D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1 \right)^2 \right\} \right] (R-r) \\ &\quad \leq \frac{1}{2}(R-r). \end{aligned}$$

Our final example relates to triangular discrimination, which arises with $f(u) = (u-1)^2/(u+1)$. We have

$$f(u) = u + 1 - \frac{4u}{u+1}, \quad f^{'}(u) = 1 - \frac{4}{(u+1)^2}$$

for $u \in [0,\infty)$, so that

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

Also

$$\frac{1}{R-r}\int_r^R f(u)du = \frac{R+r}{2} + \ln\left(\frac{R+1}{r+1}\right)^{4/(R-r)} - 3$$

and Theorem 1 provides

$$\begin{aligned}
&\left|\Delta(p,q) - \left[\frac{R+r}{2} + \ln\left(\frac{R+1}{r+1}\right)^{4/(R-r)} - 3\right]\right| \\
&\quad \le \left[\frac{1}{4} + \frac{1}{(R-r)^2}\left\{D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1\right)^2\right\}\right]\frac{(R-r)(R-1)(R+3)}{(R+1)^2} \\
&\quad \le \frac{1}{2}\frac{(R-r)(R-1)(R+3)}{(R+1)^2}.
\end{aligned}$$

4. Some numerical examples

One situation of practical interest is where p_i and q_i are close, so that we have $p_i = p_i(\varepsilon)$, $q_i = q_i(\varepsilon)$ and

$$\left|\frac{p_i(\varepsilon)}{q_i(\varepsilon)} - 1\right| \le \varepsilon \ \varepsilon \in (0,1) \tag{4.1}$$

for all $i \in \{1,...,n\}$. With $R = \varepsilon + 1$ and $r = 1 - \varepsilon$, we obtain from (3.1) that

$$\begin{aligned}
&\left|D(p(\varepsilon),q(\varepsilon)) - \frac{1}{2}\ln\left[I\left((1+\varepsilon)^2,(1-\varepsilon)^2\right)\right]\right| \\
&\quad \le \frac{\varepsilon}{2}\left[1 + \frac{1}{\varepsilon^2}D_{\chi^2}(p(\varepsilon),q(\varepsilon))\right]\ln\left[e(1+\varepsilon)\right] \\
&\quad \le \varepsilon\ln\left[e(1+\varepsilon)\right].
\end{aligned}$$

Consequently if $p(\varepsilon)$, $q(\varepsilon)$ are in the sense of (4.1), we can approximate the Kullback-Leibler distance $D(p(\varepsilon),q(\varepsilon))$ by $(1/2)\ln\left[I\left((1+\varepsilon)^2,(1-\varepsilon)^2\right)\right]$ and the error of the approximation is less than

$$E(\varepsilon) := \varepsilon\ln[e(1+\varepsilon)].$$

From (3.2), we derive

$$\begin{aligned}
\left|D(q(\varepsilon),p(\varepsilon)) - \ln\left[\frac{1}{I(1-\varepsilon,1+\varepsilon)}\right]\right| &\le \frac{1}{2}\frac{\varepsilon}{1-\varepsilon}\left[1 + \frac{1}{\varepsilon^2}D_{\chi^2}(p(\varepsilon),q(\varepsilon))\right] \\
&\le \frac{\varepsilon}{1-\varepsilon}
\end{aligned}$$

for $\varepsilon \in (0,1)$.

Consequently for $p(\varepsilon), q(\varepsilon)$ satisfying (4.1), we can approximate the Kullback–Leibler distance $D(p(\varepsilon),q(\varepsilon))$ by $\ln\left[I^{-1}(1-\varepsilon,1+\varepsilon)\right]$ and the error of the approximation is less than $\varepsilon/(1-\varepsilon)$ for $\varepsilon \in (0,1)$.

5. Application to mutual information

We consider mutual information, which is a measure of the amount of information that one random variable provides about another. It is the reduction of uncertainty about one variable due to knowledge of the other (see, for example, [1]).

Definition 1. *Consider two discrete–valued 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,*

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

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

We assume that

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

Much as with r, R we have $s \le 1 \le S$.

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

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

Inequality (3.1) yields the following proposition.

Proposition 2. *If t, p and q satisfy (5.1), then*

$$\begin{aligned} &\left| I(X;Y) - \frac{s+S}{4} \ln\left[I\left(s^2, S^2\right)\right] \right| \\ &\le \left[\frac{1}{4} + \frac{1}{(S-s)^2} \left[I_{\chi^2}(X;Y) + \left(\frac{s+S}{2} - 1 \right)^2 \right] \right] (S-s)\ln[eS] \\ &\le \frac{1}{2}(S-s)\ln(eS). \end{aligned}$$

Remark 1. *The condition $t(x,y) = p(x)p(y)$ for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$ means that the random variables X and Y are independent. We may refer to them as "quasi–independent" if*

$$\left| \frac{t(x,y)}{p(x)q(y)} - 1 \right| \le \delta \qquad (\delta \in (0,1))$$

for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$. When this occurs, we can approximate the mutual information $I(X;Y)$ by

$$\frac{1}{2}\left[\frac{(1+\delta)^2 \ln(1+\delta)^2 - (1-\delta)^2 \ln(1-\delta)^2}{4\delta} - 1 \right] \qquad (\delta \in (0,1))$$

with an error less than $E(\delta)$ for $t \in (0,1)$.

References

[1] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, 1991.

[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] D. Dacunha–Castelle, *Ecole d'été de probabilités de Saint–Fleur*, VII-1977, Berlin, Heidelberg, New York: Springer, 1978.

[6] 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.

[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] L. LeCam, *Asymptotic methods in statistical decision theory*, New York: Springer, 1986.

[9] D. S. Mitrinović, *Analytic inequalities*, Springer Verlag, 1970.

[10] F. Topsoe, Some inequalities for information divergence and related measures of discrimination, *RGMIA Res. Rep. Coll.* **2** (1999) No. 1, 85–98.

Department of Computer and Mathematical Sciences, Victoria University of Technology, PO Box 14428, MCMC, Melbourne, Victoria 8001, Australia
E-mail address: sever@matilda.vu.edu.au

Royal Australian Air Force, ARDU, P.O. 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