THE APPROXIMATION OF CSISZÁR f–DIVERGENCE FOR MAPPINGS OF BOUNDED VARIATION

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

Abstract. We consider the approximation of the Csiszár f–divergence when f or its first derivative is a function of bounded variation. The approximants are suggested by numerical integration theory.

1. Introduction

In this article we complete a chain of ideas developed in three companion papers [5]–[7] concerning measures of the divergence between two probability measures $p = (p_1, \ldots, p_n)$, $q = (q_1, \ldots, q_n)$ defined over a common set of events. Different applications have provoked different measures. A wide spectrum of measures in use has been subsumed by Csiszár [1]–[3]: the Csiszár f–divergence between p and q is defined by the functional

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

Two important instances, which we shall invoke shortly, are the variational distance $V(p,q)$ and the chi–squared divergence $D_{\chi^2}(p,q)$, for which

$$f(u) = |u-1|^m$$

with $m = 1, 2$ respectively. We address the situation in which there exist constants r, R with

$$0 < r < 1 < R < \infty \text{ and } r \le p_i/q_i \le R \text{ for } i = 1, \ldots, n. \tag{1.1}$$

Our initial work [5] assumed the individual values p_i, q_i are not known. The following result was derived.

Theorem A. *Suppose $f : [r,R] \to \mathbf{R}$ is absolutely continuous and that the derivative $f' : [r,R] \to R$ is essentially bounded, that is, $f' \in L_\infty[r,R]$. Then under the condition (1.1) we have*

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

1991 *Mathematics Subject Classification.* 94A17, 26D15.
Key words and phrases. Csiszár f–divergence, Ostrowski's inequality, bounded variation.

The two companion papers in this volume consider the approximation question: how closeness of f and g is reflected in closeness of I_f and I_g. Suppose

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

The following was derived in [6].

Theorem B. *Suppose $f : [r,R] \to \mathbf{R}$ with f' absolutely continuous on $[r,R]$ and $f'' \in L_\infty[r,R]$. If (1.1) holds, then*

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

Similarly define

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

The following was shown in [7].

Theorem C. *Suppose $f : [r,R] \to \mathbf{R}$ with f' absolutely continuous on $[r,R]$ and $f'' \in L_\infty[r,R]$. If (1.1) applies, then*

$$\begin{aligned}\left|I_f(p,q) - I_{f^\dagger}(p,q)\right| &\le \frac{1}{4}\left\|f''\right\|_\infty D_{\chi^2}(p,q) - \frac{1}{4\left\|f''\right\|_\infty} I_{f_0}(p,q)\\ &\le \frac{1}{4}\left\|f''\right\|_\infty D_{\chi^2}(p,q)\\ &\le \frac{1}{4}\left\|f''\right\|_\infty (R-1)(1-r)\\ &\le \frac{1}{16}\left\|f''\right\|_\infty (R-r)^2,\end{aligned}$$

where

$$f_0(u) := \left[f'(u) - f'(1)\right]^2.$$

These and further results in [6] and [7] assume the absolute continuity of f' (or higher derivative) and essential boundedness of f'' (or higher derivative). In this concluding contribution we weaken these assumptions to the bounded variation of f or f' and derive analogues of Theorems A, B, C. We shall employ *inter alia* the following useful lemma, which is a special case of Proposition 1 of [6].

Lemma A. *If (1.1) is satisfied, then*

$$V(p,q) \le \frac{2(R-1)(1-r)}{R-r} \le \frac{R-r}{2}.$$

The first inequality is an equality if and only if 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 $R + r = 2$.

In Section 2 we give a basic Ostrowski–type inequality for functions of bounded variation and use this to obtain a trapezoidal inequality for functions of bounded variation. In Section 3 we deduce analogues of Theorems A–C and in Section 4 we give two illustrative examples.

2. Preliminaries

We start with the following proposition which provides an Ostrowski–type inequality for mappings of bounded variation. This has been established by one of the present authors in the preprint [4]. We give a simple proof.

Proposition 1. *Suppose $g : [a,b] \to \mathbf{R}$ is of bounded variation on $[r,R]$. Then for all $x \in [a,b]$,*

$$\left| \int_a^b g(t)dt - g(x)(b-a) \right| \le \left[\frac{b-a}{2} + \left| x - \frac{a+b}{2} \right| \right] \bigvee_a^b (g) \le (b-a) \bigvee_a^b (g), \tag{2.1}$$

where $\bigvee_a^b (g)$ denotes the total variation of g on $[a,b]$. The constant $1/2$ is best–possible.

Proof. Using the integration by parts formula for a Riemann–Stieltjes integral, we have that $\int_a^x (t-a)dg(t)$ and $\int_x^b (t-b)dg(t)$ exist and that

$$\int_a^x (t-a)dg(t) = g(x)(x-a) - \int_a^x g(t)dt$$

and

$$\int_x^b (t-b)dg(t) = g(x)(b-x) - \int_x^b g(t)dt$$

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

$$g(x)(b-a) - \int_a^b g(t)dt = \int_a^x (t-a)dg(t) + \int_x^b (t-b)dg(t)$$

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

Now if $p, g : [a,b] \to \mathbf{R}$ with p continuous and g of bounded variation, then $\int_a^b p(t)dg(t)$ exists and

$$\left| \int_a^b p(t)dg(t) \right| = \sup_{t \in [a,b]} |p(t)| \bigvee_a^b (g).$$

Hence

$$\begin{aligned}
\left| g(x)(b-a) - \int_a^b g(t)dt \right| &= \left| \int_a^x (t-a)dg(t) + \int_x^b (t-b)dg(t) \right| \\
&\le \left| \int_a^x (t-a)dg(t) \right| + \left| \int_x^b (t-b)dg(t) \right| \\
&\le \sup_{t\in[a,x]} |t-a| \bigvee_a^x (g) + \sup_{t\in[x,b]} |t-b| \bigvee_x^b (g) \\
&= (x-a)\bigvee_a^x (g) + (b-x)\bigvee_x^b (g) \\
&\le \max\{x-a, b-x\} \left[\bigvee_a^x (g) + \bigvee_x^b (g) \right] \\
&= \left[\frac{b-a}{2} + \left| x - \frac{a+b}{2} \right| \right] \bigvee_a^b (g),
\end{aligned}$$

and the first inequality in (2.1) is proved. The second follows, since $|x-(a+b)/2| \le (b-a)/2$.

Suppose that (2.1) holds with a constant $c > 0$, that is,

$$\left| \int_a^b g(t)dt - g(x)(b-a) \right| \le \left[c(b-a) + \left| x - \frac{a+b}{2} \right| \right] \bigvee_a^b (g) \tag{2.2}$$

for all $x \in [a,b]$, and define $g_1 : [a,b] \to \mathbf{R}$ by

$$g_1(x) = \begin{cases} 1 & \text{if } x = (a+b)/2 \\ 0 & \text{otherwise.} \end{cases}$$

Then g_1 is of bounded variation on $[a,b]$ and

$$\bigvee_a^b (g_1) = 2, \qquad \int_a^b g_1(t)dt = 0.$$

Put $g = g_1$ and $x = (a+b)/2$ in (2.2). Then we get $1 \le 2c$, which shows that $c = 1/2$ is best–possible. □

Proposition 2. *If $g : [a,b] \to \mathbf{R}$ is of bounded variation, then for all $x_1, x_2 \in [a,b]$,*

$$\left| \int_a^b g(t)dt - \frac{b-a}{2} \sum_{i=1}^2 g(x_i) \right| \le \left[\frac{b-a}{2} + \frac{1}{2} \sum_{i=1}^2 \left| x_i - \frac{a+b}{2} \right| \right] \bigvee_a^b (g). \tag{2.3}$$

Proof. This follows by putting $x = x_i$ in Proposition 1, summing over i and then using the triangle inequality. □

Lemma 1. *Suppose $f : [r, R] \to \mathbf{R}$ is differentiable, so that $f^{'}$ is of bounded variation on $[r, R]$. If $r < 1 < R$ and $x \in [r, R]$, then*

$$\left| f(x) - f(1) - \frac{x-1}{2}\left[f^{'}(1) + f^{'}(x) \right] \right| \leq |x-1| \bigvee_{r}^{R}(f^{'}). \tag{2.4}$$

Proof. For $x \geq 1$, we set $g = f^{'}$, $x_1 = a = 1$ and $x_2 = b = x$ in (2.3) to derive

$$\left| f(x) - f(1) - \frac{x-1}{2}\left[f^{'}(1) + f^{'}(x) \right] \right| \leq (x-1) \bigvee_{1}^{x}(f^{'}) \leq (x-1) \bigvee_{r}^{R}(f^{'}).$$

Similarly if $x < 1$, we set $g = f^{'}$, $x_1 = a = x$ and $x_2 = b = 1$ in (2.3) to derive

$$\left| f(1) - f(x) - \frac{1-x}{2}\left[f^{'}(1) + f^{'}(x) \right] \right| \leq (1-x) \bigvee_{x}^{1}(f^{'}) \leq (1-x) \bigvee_{r}^{R}(f^{'}).$$

The desired result follows in both cases. □

3. Main results

Theorem 1. *If $f : [r, R] \to \mathbf{R}$ is of bounded variation and (1.1) applies, then*

$$\begin{aligned} \left| I_f(p,q) - \frac{1}{R-r} \int_{r}^{R} f(t)dt \right| & \\ \leq & \left[\frac{1}{2} + \frac{1}{R-r} \sum_{k=1}^{n} \left| p_i - \frac{r+R}{2} \cdot q_i \right| \right] \bigvee_{r}^{R}(f) \\ \leq & \left\{ \frac{1}{2} + \frac{1}{R-r}\left[V(p,q) + \left| \frac{r+R}{2} - 1 \right| \right] \right\} \bigvee_{r}^{R}(f) \\ \leq & \left[1 + \frac{1}{R-r} \cdot V(p,q) \right] \bigvee_{r}^{R}(f) \\ \leq & \frac{3}{2} \bigvee_{r}^{R}(f). \end{aligned} \tag{3.1}$$

Proof. The choices $g = f$, $x = p_i/q_i$ $(i = 1, ..., n)$, $a = r$, $b = R$ in (2.1) give

$$\left| f\left(\frac{p_i}{q_i}\right) - \frac{1}{R-r}\int_{r}^{R} f(t)dt \right| \leq \left[\frac{1}{2} + \frac{1}{R-r}\left| \frac{p_i}{q_i} - \frac{r+R}{2} \right| \right] \bigvee_{r}^{R}(f)$$

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

If we multiply by q_i and sum over i, we obtain *via* the generalized triangle inequality that

$$\left| I_f(p,q) - \frac{1}{R-r}\int_{r}^{R} f(t)dt \right| \leq \left[\frac{1}{2} + \frac{1}{R-r}\sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - \frac{r+R}{2} \right| \right] \bigvee_{r}^{R}(f),$$

whence we have the first inequality in (3.1).

The second follows from

$$\sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - 1 - \left(\frac{r+R}{2} - 1 \right) \right| \leq \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - 1 \right| + \left| \frac{r+R}{2} - 1 \right|$$

$$= V(p,q) + \left| \frac{r+R}{2} - 1 \right|$$

and the third from

$$\left| \frac{r+R}{2} - 1 \right| \leq \frac{R-r}{2}.$$

The final inequality follows by Lemma A. □

The following corollary emphasizes better the approximation aspect of the theorem.

Corollary 1. *Let $f : [0,2] \to \mathbf{R}$ be a mapping of bounded variation. If $\eta \in (0,1)$ and $p(\eta)$ and $q(\eta)$ are probability distributions satisfying*

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

then

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

and the reminder term R_f satisfies

$$|R_f(p,q,\eta)| \leq \frac{1}{2} \left[1 + \frac{1}{\eta} V(p(\eta), q(\eta)) \right] \bigvee_{1-\eta}^{1+\eta} (f).$$

We note that the best inequality we can get from (2.1) is

$$\left| \int_a^b g(t)dt - g\left(\frac{a+b}{2} \right)(b-a) \right| \leq \frac{b-a}{2} \bigvee_a^b (g), \tag{3.2}$$

which arises for $x = (a+b)/2$.

We now turn to a comparison theorem which encapsulates bounded–variation analogues of Theorems B and C.

Theorem 2. *Suppose $f : [r,R] \to \mathbf{R}$ is differentiable and so f' is of bounded variation. If (1.1) applies, then*

$$|I_f(p,q) - I_{f^*}(p,q)| \leq \frac{1}{2} V(p,q) \bigvee_r^R (f') \leq \frac{R-1}{4} \bigvee_r^R (f'), \tag{3.3}$$

$$\left| I_f(p,q) - I_{f^\dagger}(p,q) \right| \leq V(p,q) \bigvee_r^R (f') \leq \frac{R-1}{2} \bigvee_r^R (f'). \tag{3.4}$$

Proof. Taking (3.2) with $g = f^{'}$, $a = 1$, and $b = x \in [r, R]$ gives

$$|f(x) - f^*(x)| \leq \frac{|x-1|}{2} \bigvee_1^x (f^{'}) \leq \frac{|x-1|}{2} \bigvee_r^R (f^{'})$$

for all $x \in [r, R]$. The first inequality in (3.3) follows by putting $x = p_i/q_i$, multiplying by q_i, summing over $i = 1, \ldots, n$ and using the generalized triangle inequality. The second inequality is given by Lemma A.

The proof of (3.4) follows similarly from (2.4). □

Both parts may be viewed in terms of approximation. Thus for (3.3) we have the following.

Corollary 2. *Suppose $f : [0, 2] \to \mathbf{R}$ has its first derivative of bounded variation. If $\eta \in (0, 1)$ and $p(\eta)$, $q(\eta)$ are probability distributions satisfying,*

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

then

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

and the reminder R_f is such that

$$|R_f(p, q, \eta)| \leq \frac{1}{2} V(p(\eta), q(\eta)) \bigvee_{1-\eta}^{1+\eta} (f^{'}) \leq \frac{\eta}{2} \bigvee_{1-\eta}^{1+\eta} (f^{'}).$$

4. Examples

Suppose (1.1) holds and $f : [r, R] \to \mathbf{R}$ is given by $f(u) = u \ln u$, so that $I_f(p, q)$ is the Kullback–Leibler distance

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

We have

$$\begin{aligned} \int_r^R f(t)dt &= \frac{1}{4} \left[R^2 \ln R^2 - r^2 \ln r^2 - \left(R^2 - r^2 \right) \right] \\ &= \frac{R^2 - r^2}{4} \ln \left[I \left(R^2, r^2 \right) \right], \end{aligned}$$

where $I(a, b)$ is the identric mean of two positive numbers a, b and is given by

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

Also

$$\bigvee_r^R (f) = \int_a^b \left| f^{'}(t) \right| dt = \int_a^b |\ln(et)| \, dt =: \lambda(r, R).$$

If $0 < r \leq 1/e$, then

$$\lambda(r,R) = \int_r^{1/e} [-\ln(et)]\,dt + \int_{1/e}^{R} \ln(et)dt$$
$$= -\frac{1}{e}\int_r^{1/e} \ln(et)d(et) + \frac{1}{e}\int_{1/e}^{R} \ln(et)d(et).$$

Since

$$\int_\alpha^\beta \ln x \; dx = \ln I(\alpha,\beta) \text{ for } \alpha,\beta > 0,$$

we have

$$\lambda(r,R) = -\frac{1}{e}\ln\left[I\left(r,e^{-1}\right)\right] + \frac{1}{e}\ln\left[I(e^{-1},R)\right] = \ln\left[\frac{I(e^{-1},R)}{I(r,e^{-1})}\right]^{1/e}.$$

If on the other hand $1/e < r < 1$, then

$$\lambda(r,R) = \int_r^R \ln(et)dt = \frac{1}{e}\ln I(r,R) = \ln\left[I(r,R)\right]^{1/e}.$$

Thus (3.1) gives

$$\begin{aligned}
&\left|D(p,q) - \frac{R+r}{4}\ln\left[I\left(R^2,r^2\right)\right]\right| \\
&\leq \left[\frac{1}{2} + \frac{1}{R-r}\sum_{i=1}^n \left|p_i - \frac{r+R}{2}\cdot q_i\right|\right]\lambda(r,R) \\
&\leq \left\{\frac{1}{2} + \frac{1}{R-r}\left[V(p,q) + \left|\frac{r+R}{2} - 1\right|\right]\right\}\lambda(r,R) \\
&\leq \left[1 + \frac{1}{R-r}V(p,q)\right]\lambda(r,R) \\
&\leq \frac{3}{2}\lambda(r,R).
\end{aligned}$$

Also

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

and

$$\bigvee_r^R (f') = \int_r^R \left|f''(t)\right| dt = \int_r^R \frac{dt}{t} = \ln\left(\frac{R}{r}\right).$$

Hence by (3.3) we have

$$\begin{aligned}\left|D(p,q)-\sum_{i=1}^{n}(p_i-q_i)\ln\left(\frac{p_i+q_i}{2q_i}\right)\right| &\le \frac{1}{2}V(p,q)\ln\left(\frac{R}{r}\right)\\ &\le \frac{R-r}{4}(\ln R-\ln r)\\ &= \frac{(R-r)^2}{4L(r,R)},\end{aligned}$$

where $L(a,b)$ is the logarithmic mean which for positive arguments a, b is given by

$$L(a,b):=\begin{cases} a & \text{if } b=a\\ \frac{b-a}{\ln b-\ln a} & \text{otherwise.}\end{cases}$$

Therefore we have the inequality

$$\left|D(p,q)-\sum_{i=1}^{n}(p_i-q_i)\ln\left(\frac{p_i+q_i}{2q_i}\right)\right| \le \frac{1}{4}\cdot\frac{R-r}{L(r,R)}V(p,q)\le\frac{(R-r)^2}{4L(r,R)}.$$

Finally suppose $f:[r,R]\to\mathbf{R}$ is given by $f(u)=|u-1|$, so $I_f(p,q)$ becomes the variational distance

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

We have

$$\begin{aligned}\frac{1}{R-r}\int_r^R f(t)dt &= \frac{1}{R-r}\int_r^R|u-1|\,du\\ &= \frac{1}{R-r}\left[\int_r^1(1-u)du+\frac{1}{R-r}\int_1^R(u-1)\,du\right]\\ &= \frac{1}{R-r}\left[\frac{(r-1)^2}{2}+\frac{(R-1)^2}{2}\right]\\ &= \frac{1}{R-r}\left[\frac{(R-r)^2}{4}+\left(\frac{r+R}{2}-1\right)^2\right]\end{aligned}$$

and

$$\bigvee_r^R(f)=\bigvee_r^1(f)+\bigvee_1^R(f)=1-r+R-1=R-r.$$

By (3.1) we have

$$\begin{aligned}
&\left|V(p,q) - \frac{1}{R-r}\left[\frac{(R-r)^2}{4} + \left(\frac{r+R}{2} - 1\right)^2\right]\right| \\
&\leq \left[\frac{1}{2} + \frac{1}{R-r}\sum_{i=1}^{n}\left|p_i - \frac{r+R}{2}\cdot q_i\right|\right](R-r) \\
&\leq \left\{\frac{1}{2} + \frac{1}{R-r}\left[V(p,q) + \left|\frac{r+R}{2} - 1\right|\right]\right\}(R-r) \\
&\leq \left[1 + \frac{1}{R-r}V(p,q)\right](R-r) \\
&\leq \frac{3(R-r)}{2}.
\end{aligned}$$

References

[1] I. Csiszár, *A note on Jensen's inequality*, Studia Sci. Math. Hung. **1** (1966), 185–188.
[2] I. Csiszár, *Information–type measures of differences of probability distributions and indirect observation*, Studia Sci. Math. Hung. **2** (1967), 299–318.
[3] I. Csiszár, *On topological properties of f–divergence*, Studia Sci. Math. Hung. **2** (1967), 329–339.
[4] S. S. Dragomir, *On the Ostrowski's integral inequality for mappings of bounded variation and applications*, RGMIA Res. Rep. Coll. **2** (1999) No. 1, 73–80.
[5] S. S. Dragomir, V. Gluščević & C. E. M. Pearce, *Csiszár f–divergence, Ostrowski's inequality and mutual information*, to appear in J. Nonlin. Anal. Ser. A.
[6] S. S. Dragomir, V. Glusčević & C. E. M. Pearce, *Approximations for Csiszár f–divergence via midpoint inequalities*, submitted.
[7] S. S. Dragomir, V. Gluščević & C. E. M. Pearce, *The approximations for Csiszár f–divergence for absolutely continuous mappings*, submitted.
[8] S. S. Dragomir & 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.

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

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