On the Approximation of the Integral Mean Divergence and f-Divergence via Mean Results

ON THE APPROXIMATION OF THE INTEGRAL MEAN DIVERGENCE AND $f-$DIVERGENCE VIA MEAN RESULTS

P. CERONE AND S.S. DRAGOMIR

Abstract. Results involving the approximation of the difference between two integral means are utilised to obtain bounds on the Integral Mean Divergence and the $f-$divergence due to Csiszár. The current work does not restrict the functions involved to be convex. If convexity is imposed then the Integral Mean Divergence is the Hermite-Hadamard divergence introduced by Shioya and Da-te.

1. Introduction

A plethora of divergence measures have been introduced in the literature in an effort to tackle an important issue of many applications of probability theory. Namely, that of an appropriate measure of difference or distance or discrimination between two probability distributions. These measures have been applied in a variety of fields including: anthropology, genetics, finance, biology, signal processing, pattern recognition, approximation of probability distributions, computational learning and so on.

The reader is referred to the paper by Kapur [17] and the book online [18] by Taneja for an extensive presentation of various divergence measures. Many, although not all, are special instances of Csiszár's $f-$divergence [1] – [3], $D_f(p,q)$. Assume that for a given set χ and a $\sigma-$finite measure μ, the set of all probability density functions on μ is

$$\Omega := \left\{ p | p : \chi \to \mathbb{R},\ p(x) \geq 0,\ \int_\chi p(x)\, d\mu(x) = 1 \right\}. \tag{1.1}$$

The $f-$divergence introduced by Csiszár [3] is then defined by

$$D_f(p,q) := \int_\chi p(x) f\left(\frac{q(x)}{p(x)}\right) d\mu(x), \quad p,q \in \Omega, \tag{1.2}$$

where f is assumed convex on $(0,\infty)$. It is further commonly imposed that $f(u)$ be zero and strictly convex at $u=1$. Shioya et al. [14] present three basic properties of $D_f(p,q)$ as:

Date: November 07, 2001.

1991 *Mathematics Subject Classification.* Primary 94A17; Secondary 26D15.

Key words and phrases. Integral Mean Divergence, Csiszár's $f-$Divergence, Approximation, Hermite-Hadamard Divergence.

$(p.1)$	*Non-negativity:*	$D_f(p,q) \geq 0,$	with equality if and only if $p \equiv q$ on χ.
$(p.2)$	*Duality:*	$D_f(p,q) = D_{f^*}(q,p)\,,$	$f^*(u) = uf\left(\frac{1}{u}\right)$
$(p.3)$	*Invariance:*	$D_f(p,q) = D_{f^\dagger}(p,q)\,,$	$f^\dagger(u) = f(u) + c(u-1)\,, \quad c \in \mathbb{R}.$

For f convex on $(0,\infty)$ and $f(u) = 0$ and strictly convex at $u = 1$ then $(p.1)$ holds. It may still hold if f is not restricted to these convexity properties. Properties $(p.2)$ and $(p.3)$ hold for any f with $(p.3)$ relying on the fact that $\int_\chi (q(x) - p(x))\, d\mu(x) = 0$ since $p, q \in \Omega$.

For g convex, Shioya and Da-te [13, 14] introduced the Hermite-Hadamard divergence measure

$$D_{HH}^g(p,q) := \int_\chi p(x) \frac{\int_1^{\frac{q(x)}{p(x)}} g(t)\, dt}{\frac{q(x)}{p(x)} - 1} d\mu(x)\,. \tag{1.3}$$

and showed that properties $(p.1)$ – $(p.3)$ also hold for $D_{HH}^g(p,q)$.

By the use of the Hermite-Hadamard inequality, they also proved that, for g a normalised convex function, the following inequality holds:

$$D_g\left(p, \frac{1}{2}p + \frac{1}{2}q\right) \leq D_{HH}^g(p,q) \leq \frac{1}{2} D_g(p,q)\,. \tag{1.4}$$

2. The Integral Mean Divergence

The current paper will make use of the following result by the authors and coworkers, [10] providing an estimate of the difference between integral means.

Theorem 1. *Let $g : [a,b] \to \mathbb{R}$ be an absolutely continuous mapping with $g' \in L_\infty[a,b]$ so that $\|g'\|_\infty := ess \sup_{t\in[a,b]} |g'(t)| < \infty$ then for $a \leq c < d \leq b$ the inequalities*

$$\begin{aligned} &\left| \frac{1}{b-a}\int_a^b g(t)\, dt - \frac{1}{d-c}\int_c^d g(t)\, dt \right| \\ \leq\ & \left\{ \frac{1}{4} + \left(\frac{\frac{a+b}{2} - \frac{c+d}{2}}{(b-a)-(d-c)} \right)^2 \right\} [b-a-(d-c)] \|g'\|_\infty \\ \leq\ & \frac{1}{2}[(b-a)-(d-c)] \|g'\|_\infty\,, \end{aligned} \tag{2.1}$$

hold.

The constants $\frac{1}{4}$ and $\frac{1}{2}$ in the above inequalities are best possible.

Remark 1. *It should be noted that if we define the integral mean of a function $g(\cdot)$ over the interval $[a,b]$ by*

$$\mathcal{M}(g;a,b) := \begin{cases} \frac{1}{b-a}\int_a^b g(t)\, dt, & b \neq a \\ g(a)\,, & b = a \end{cases} \tag{2.2}$$

then the results of Theorem 1 may be written as

$$\begin{aligned} (2.3) \qquad \left|\mathcal{M}(g;a,b)-\mathcal{M}(g;c,d)\right| &\leq C(a,c,d,b)\left\|g'\right\|_\infty \\ &\leq \frac{1}{2}\left[(b-d)+(c-a)\right]\left\|g'\right\|_\infty, \end{aligned}$$

where

$$(2.4) \qquad C(a,c,d,b):=\frac{1}{2}\left[\frac{(b-d)^2+(c-a)^2}{b-d+c-a}\right].$$

The second inequality in (2.3) is obvious from the first on noting that $A^2+B^2\leq(A+B)^2$, *for* $A,B>0$ *and using (2.4).*

It should finally be noted that even if the requirement for $c<d$ in Theorem 1 was omitted, the results would still hold. The requirement was made for definiteness.

We define the *integral mean divergence measure*

$$(2.5) \qquad D_{\mathfrak{M}(g)}(p,q):=\int_\chi p(x)\,\mathfrak{M}(g)\left(\frac{q(x)}{p(x)}\right)d\mu(x),$$

where, from (2.2)

$$(2.6) \qquad \mathfrak{M}(g)(z):=\mathcal{M}(g;1,z)=\begin{cases}\frac{1}{z-1}\int_1^z g(u)\,du, & z\neq 1\\ g(1), & z=1\end{cases}.$$

We note that if $g(\cdot)$ is convex, then (2.5) – (2.6) is the Hermite-Hadamard divergence measure $D_{HH}^g(p,q)$ defined by (1.3), so named by Shioya and Da-te [13] since they utilised the well known Hermite-Hadamard inequality for convex functions to procure bounds. They showed that

$$(2.7) \qquad D_g\left(p,\frac{p+q}{2}\right)\leq D_{HH}^g(p,q)\leq\frac{1}{2}D_g(p,q),$$

where the lower bound is the generalised Lin-Wong $g-$divergence and the upper bound is one half of Csiszár's divergence, (1.2). In (2.7) $g(\cdot)$ is both convex and normalised so that $g(1)=0$.

The following theorem produces bounds for the integral mean divergence measure $D_{\mathfrak{M}(g)}(p,q)$ defined by (2.5) – (2.6) where $g(\cdot)$ is not assumed to be convex.

Theorem 2. *Let* $g:\mathbb{R}\to\mathbb{R}$ *be absolutely continuous on any* $[a,b]\subset\mathbb{R}$. *If* $p,q\in\Omega$ *then with* $0\leq r\leq 1\leq R$ *and* $r\leq\frac{q(x)}{p(x)}\leq R<\infty$ *for* $x\in\chi$ *we have*

$$\begin{aligned} (2.8) \qquad & \left|D_{\mathfrak{M}(g)}(p,q)-\mathcal{M}(g;r,R)\right| \\ \leq\ & \left\|g'\right\|_\infty\left\{\frac{R-r}{2}-(1-r)\int_\chi\frac{p(x)\left[Rp(x)-q(x)\right]}{(R+1-r)\,p(x)-q(x)}d\mu(x)\right\} \\ \leq\ & \left[\frac{R-r}{2}-(R-1)\left(\frac{1-r}{R+1-r}\right)^2\right]\left\|g'\right\|_\infty \\ \leq\ & \frac{R-r}{2}\left\|g'\right\|_\infty. \end{aligned}$$

Proof. Let $a = r$, $b = R$ and $c = 1$, $d = z$ then from (2.3) we have

$$\begin{aligned}\left|\mathcal{M}\left(g;1,z\right)-\mathcal{M}\left(g;r,R\right)\right| &\leq C\left(r,1,z,R\right)\left\|g'\right\|_{\infty}\\ &\leq \left(R-z+1-r\right)\frac{\left\|g'\right\|_{\infty}}{2}\end{aligned}$$

which, upon choosing $z = \frac{q(x)}{p(x)}$ for $x \in \chi$ we obtain

$$\begin{aligned}\left|\mathcal{M}\left(g;1,\frac{q\left(x\right)}{p\left(x\right)}\right)-\mathcal{M}\left(g;r,R\right)\right| &\leq C\left(r,1,\frac{q\left(x\right)}{p\left(x\right)},R\right)\left\|g'\right\|_{\infty}\\ &\leq \left[\left(R-r+1\right)p\left(x\right)-q\left(x\right)\right]\frac{\left\|g'\right\|_{\infty}}{2},\end{aligned} \tag{2.9}$$

where $C\left(a,c,d,b\right)$ is as defined in (2.4).

On multiplication of (2.9) by $p\left(x\right) \geq 0$ for all $x \in \chi$ and integration with respect to the measure μ we obtain the first and third inequalities of (2.8). Here we have used the fact that

$$p\left(x\right)C\left(r,1,\frac{q\left(x\right)}{p\left(x\right)},R\right)=\frac{A^{2}+B^{2}}{2\left(A+B\right)}=\frac{A+B}{2}-\frac{AB}{A+B} \tag{2.10}$$

with $A = Rp\left(x\right) - q\left(x\right)$ and $B = \left(1-r\right)p\left(x\right)$.

Now, to obtain the second inequality in (2.8) we have

$$\begin{aligned}\frac{p\left(x\right)\left[Rp\left(x\right)-q\left(x\right)\right]}{\left(R+1-r\right)p\left(x\right)-q\left(x\right)} &= \frac{R}{R+1-r}p\left(x\right)-\frac{1-r}{\left(R+1-r\right)^{2}}q\left(x\right)\\ &\quad -\frac{1-r}{\left(R+1-r\right)^{2}}\cdot\frac{q^{2}\left(x\right)}{\left(R+1-r\right)p\left(x\right)-q\left(x\right)}.\end{aligned}$$

Hence

$$\begin{aligned}&-\left(1-r\right)\int_{\chi}\frac{p\left(x\right)\left[Rp\left(x\right)-q\left(x\right)\right]}{\left(R+1-r\right)p\left(x\right)-q\left(x\right)}d\mu\left(x\right)\\ &= \frac{-R\left(1-r\right)}{R+1-r}+\left(\frac{1-r}{R+1-r}\right)^{2}\\ &\quad+\left(\frac{1-r}{R+1-r}\right)^{2}\int_{\chi}\frac{q^{2}\left(x\right)}{\left(R+1-r\right)p\left(x\right)-q\left(x\right)}d\mu\left(x\right).\end{aligned} \tag{2.11}$$

Further, since $0 < r \leq \frac{q(x)}{p(x)} \leq R$ then $\frac{1}{R} < \frac{p(x)}{q(x)} < \frac{1}{r}$ giving

$$\frac{1}{\frac{R+1-r}{r}-1}<\frac{1}{\left(R+1-r\right)\frac{p(x)}{q(x)}-1}<\frac{1}{\frac{R+1-r}{R}-1}.$$

Thus,

$$\int_{\chi}\frac{q^{2}\left(x\right)}{\left(R+1-r\right)p\left(x\right)-q\left(x\right)}d\mu\left(x\right)<\frac{R}{1-r}\int_{\chi}q\left(x\right)d\mu\left(x\right)=\frac{R}{1-r},$$

which gives from (2.11)

$$-\left(1-r\right)\int_{\chi}\frac{p\left(x\right)\left[Rp\left(x\right)-q\left(x\right)\right]}{\left(R+1-r\right)p\left(x\right)-q\left(x\right)}d\mu\left(x\right)\leq-\left(\frac{1-r}{R+1-r}\right)^{2}\left(R-1\right). \tag{2.12}$$

Substitution of (2.12) into the first inequality in (2.8) produces the second and thus the theorem is completely proved. ∎

Remark 2. *If $p(x) = q(x)$, for $x \in \chi_E$ and $p(x) \neq q(x)$ for $x \in \chi_N$ where $\chi = \chi_E \cup \chi_N$ then from (2.5) and (2.6) we have*

$$\begin{aligned} D_{\mathfrak{M}(g)}(p,q) &= g(1) \int_{\chi_E} p(x)\, d\mu(x) \\ &\quad + \int_{\chi_N} p(x) \left(\frac{1}{\frac{q(x)}{p(x)} - 1} \int_1^{\frac{q(x)}{p(x)}} g(u)\, du \right) d\mu(x). \end{aligned}$$

The degenerate case occurs if $\chi = \chi_E$.

Corollary 1. *Let the conditions of Theorem 2 hold and let $t \in [0,1]$, then the following bounds for the parametrised integral mean divergence measure $D_{\mathfrak{M}(g)}(p, tq + (1-t)p)$ hold. Namely,*

$$\begin{aligned} &\left| D_{\mathfrak{M}(g)}(p, tq + (1-t)p) - \mathcal{M}(g; r, R) \right| \\ &\leq \|g'\|_\infty \left\{ \frac{r+R}{2} - 1 + (1-r)^2 D_{f^*}(p,q) \right\} \\ &\leq \|g'\|_\infty \left\{ \frac{r+R}{2} - 1 + \frac{(1-r)^2}{R - r - t(R-1)} \right\} \\ &\leq \|g'\|_\infty \left(\frac{R-r}{2} \right), \end{aligned} \tag{2.13}$$

where $D_{\mathfrak{M}(g)}(p,q)$ is defined by (2.5) – (2.6), $\mathcal{M}(g; a, b)$ by (2.2),

$$D_{f^*}(p,q) = \int_\chi p(x) f^*\left(\frac{q(x)}{p(x)} \right) d\mu(x), \tag{2.14}$$

with

$$f^*(z) = [(R - r + t) - tz]^{-1}. \tag{2.15}$$

Proof. Substitution of $tq(x) + (1-t)p(x)$ for $q(x)$ in the first inequality in (2.8) produces

$$\begin{aligned} &\left| D_{\mathfrak{M}(g)}(p, tq + (1-t)p) - \mathcal{M}(g; r, R) \right| \\ &\leq \|g'\|_\infty \left\{ \frac{R-r}{2} - (1-r) \int_\chi p(x) \frac{[(R+t-1)p(x) - tq(x)]}{(R+t-r)p(x) - tq(x)} d\mu(x) \right\} \\ &= \|g'\|_\infty \left\{ \frac{R-r}{2} - (1-r) \int_\chi p(x)\, d\mu(x) \right. \\ &\qquad \left. + (1-r)^2 \int_\chi p(x)\, h(p,q,x;t)\, d\mu(x) \right\} \end{aligned}$$

from which the first inequality is procured, where $h(p,q,x;t)$ is given by $f^*\left(\frac{q(x)}{p(x)}\right)$ from (2.15). That is,

$$h(p,q,x;t) = \frac{1}{(R - r + t) - t\left(\frac{q(x)}{p(x)}\right)}. \tag{2.16}$$

Now, to obtain the second inequality we note that, from (2.14) – (2.16),

$$\begin{aligned} D_{f^*}(p,q) &\leq \sup_{x\in\chi}\left|f^*\left(\frac{q(x)}{p(x)}\right)\right|\int_\chi p(x)\,d\mu(x) \\ &= \sup_{x\in\chi}|h(p,q,x;t)|. \end{aligned} \tag{2.17}$$

Further, we have that $r \leq \frac{q(x)}{p(x)} \leq R$, $x\in\chi$ so that some manipulation gives

$$\frac{1}{R-r+t(1-r)} \leq h(p,q,x;t) \leq \frac{1}{R-r-t(R-1)}. \tag{2.18}$$

That is,

$$f^*(r) \leq f^*\left(\frac{q(x)}{p(x)}\right) \leq f^*(R)$$

since $f^*(u)$ is an increasing function of u.

Using (2.18) in (2.17) gives the second inequality in (2.13).

The third inequality may be obtained on noting that the second inequality in (2.13) produces the coarsest bound when $t=1$.

The corollary is now proved. ∎

Remark 3. *If g is $L-$Lipschitzian so that it satisfies $|g(x)-g(y)| \leq L|x-y|$ for all $x,y\in[a,b]$ ($[r,R]$) then the results of this section hold with $\|g'\|_\infty$ being replaced by L.*

3. The $f-$Divergence of Csiszár

The techniques of the previous section will now be applied to approximating Csiszár's $f-$divergence $D_f(p,q)$ as defined by (1.2), where here, f is not necessarily convex.

The following proposition involving the divergence measure $D_{v^m}(p,q)$ will be required, where

$$D_{v^m}(p,q) = \int_\chi p(x)\left|1-\left(\frac{q(x)}{p(x)}\right)^m\right|d\mu(x), \quad m\geq 1. \tag{3.1}$$

We note that $D_v(p,q)$ is the well-known variational distance.

Proposition 1. *Suppose that real numbers r,R exist such that $0<r\leq\frac{g(x)}{p(x)}\leq R<\infty$ for all $x\in\chi$ and $r<1<R$. Then for integer $m\geq 1$,*

$$\left(\frac{1-r^m}{1-r}\right)D_v(p,q) \leq D_{v^m}(p,q) \leq \left(\frac{R^m-1}{R-1}\right)D_v(p,q) \tag{3.2}$$

with equality holding if $m=1$. If $\frac{q(x)}{p(x)}=r$ for all $x\in\chi$ then equality is obtained for the first inequality. For $\frac{q(x)}{p(x)}=R$ for all $x\in\chi$ then the second inequality becomes an equality.

Proof. From (3.1) we have that

$$D_{v^m}(p,q) = \int_\chi p(x)\left|1-\left(\frac{q(x)}{p(x)}\right)\right|\left|\frac{1-\left(\frac{q(x)}{p(x)}\right)^m}{1-\left(\frac{q(x)}{p(x)}\right)}\right|d\mu(x). \tag{3.3}$$

Now, let $z = \frac{q(x)}{p(x)}$ then for $r \le z \le R$,

$$\frac{1-r^m}{1-r} \le \left|\frac{1-z^m}{1-z}\right| \le \frac{R^m-1}{R-1},$$

which upon substitution into (3.3) gives (3.2).

If $m = 1$ then we obtain $D_v(p,q)$. For $\frac{q(x)}{p(x)} = r$, or $\frac{q(x)}{p(x)} = R$ for all $x \in \chi$ then we have

$$(1-r^m) \le \frac{R^m-1}{R-1} \cdot (1-r).$$

∎

The following theorem holds.

Theorem 3. *Let $g : \mathbb{R} \to \mathbb{R}$ be absolutely continuous on any interval $[a,b] \subset \mathbb{R}$. If $p, q \in \Omega$, then with $0 \le r \le 1 \le R$ and $r \le \frac{q(x)}{p(x)} \le R < \infty$ for $x \in \chi$ we have, for $g(t) = f'(t)$ and $f'' \in L_\infty[r,R]$ $\left(\|f''\|_\infty := ess \sup_{t\in[r,R]} |f''(t)| < \infty\right)$,*

$$\begin{aligned} &|D_f(p,q) - f(1)| \\ \le\ & \|f''\|_\infty \left\{\frac{r+R}{2} D_v(p,q) - \frac{1}{2} D_{v^2}(p,q) \right. \\ &\left. + (1-r)^2 \int_\chi |p(x) - q(x)|\,\theta(x)\,d\mu(x)\right\} \\ \le\ & \|f''\|_\infty \left[\frac{R-r}{2} D_v(p,q) + \frac{1}{2} D_{v^2}(p,q)\right] \\ \le\ & \|f''\|_\infty \left[R + \frac{1-r}{2}\right] D_v(p,q) \end{aligned} \tag{3.4}$$

where

$$\theta(x) = \frac{p(x)}{(R+1-r)\,p(x) - q(x)}, \tag{3.5}$$

$D_{v^2}(p,q)$ is as given by (3.1) with $m = 2$, and $D_v(p,q)$ is the variational distance defined by (3.1) with $m = 1$, namely

$$D_v(p,q) := \int_\chi |p(x) - q(x)|\,d\mu(x). \tag{3.6}$$

Proof. The proof follows closely that of Theorem 2. From (2.3) we have on taking $a = r$, $b = R$ and $c = 1$, $d = z$ and so with $g(t) = f'(t)$

$$\left|\frac{f(R) - f(r)}{R-r} - \frac{f(z) - f(1)}{z-1}\right| \le C(r,1,z,R)\,\|f''\|_\infty,$$

$$\left|f(z) - f(1) - (z-1)\left(\frac{f(R) - f(r)}{R-r}\right)\right| \le |z-1|\,C(r,1,z,R)\,\|f''\|_\infty, \tag{3.7}$$

where $g(t) = f'(t)$ and $C(a,c,d,b)$ is as defined in (2.4).

Choosing $z=\frac{q(x)}{p(x)}$ for $x\in\chi$ in (3.7), multiplication of the result by $p(x)\geq 0$ and integration over χ with respect to the measure μ we obtain

$$\left|D_f(p,q)-f(1)\int_\chi p(x)\,d\mu(x)+\left(\frac{f(R)-f(r)}{R-r}\right)\int_\chi (p(x)-q(x))\,d\mu(x)\right| \leq \|f''\|_\infty\int_\chi\left|\frac{q(x)}{p(x)}-1\right|p(x)\,C\left(r,1,\frac{q(x)}{p(x)},R\right)d\mu(x). \tag{3.8}$$

We note from (2.4) and (2.10) that

$$\begin{aligned}
& p(x)\,C\left(r,1,\frac{q(x)}{p(x)},R\right)\\
&= \frac{1}{2}\cdot\frac{(Rp(x)-q(x))^2+(1-r)^2p^2(x)}{Rp(x)-q(x)+(1-r)p(x)}\\
&= \frac{(R+1-r)p(x)-q(x)}{2}-(1-r)p(x)\left(\frac{Rp(x)-q(x)}{(R+1-r)p(x)-q(x)}\right)\\
&= \frac{(R+1-r)p(x)-q(x)}{2}-(1-r)p(x)+\frac{(1-r)^2p^2(x)}{(R+1-r)p(x)-q(x)}\\
&= \frac{r+R}{2}p(x)-\frac{p(x)+q(x)}{2}+\frac{(1-r)^2p^2(x)}{(R+1-r)p(x)-q(x)},
\end{aligned}$$

giving from (3.8)

$$\begin{aligned}
|D_f(p,q)-f(1)| \leq\ & \|f''\|_\infty\left\{\frac{r+R}{2}\int_\chi|p(x)-q(x)|\,d\mu(x)\right.\\
&-\frac{1}{2}\int_\chi p(x)\left|\left(\frac{q(x)}{p(x)}\right)^2-1\right|d\mu(x)\\
&\left.+(1-r)^2\int_\chi\frac{|p(x)-q(x)|\cdot p(x)}{(R+1-r)p(x)-q(x)}d\mu(x)\right\}.
\end{aligned}$$

Here we have used the fact that $p(\cdot)$ and $q(\cdot)$ are densities with respect to the measure space μ on χ.

Now, for the second bound in (3.4). We note that since $r\leq\frac{q(x)}{p(x)}\leq R$, $x\in\chi$ then from (3.5),

$$\frac{1}{R-r+1-r}<\theta(x)<\frac{1}{1-r}.$$

Utilizing this upper bound in the first inequality of (3.4) produces the second after some simplification. In particular, we have used the fact that

$$\begin{aligned}
& D_v(p,q)-\frac{1}{2}D_{v^2}(p,q)\\
&= \int_\chi|p(x)-q(x)|\,d\mu(x)-\frac{1}{2}\int_\chi\frac{\left|q^2(x)-p^2(x)\right|}{p(x)}d\mu(x)\\
&= \int_\chi|p(x)-q(x)|\left[1-\frac{p(x)+q(x)}{2p(x)}\right]d\mu(x)\\
&= \frac{1}{2}\int_\chi p(x)\left|1-\frac{q(x)}{p(x)}\right|\left(1+\frac{q(x)}{p(x)}\right)d\mu(x).
\end{aligned}$$

We note that the second inequality in (3.4) can be obtained by using the second inequality in (2.3) directly. Finally, using the fact that

$$1+r \leq 1+\frac{q(x)}{p(x)} \leq 1+R,$$

gives the third inequality. ∎

Corollary 2. *Let the conditions of Theorem 3 persist and let $t \in [0,1]$, then the following bounds for the parametrised divergence measure $D_f(p, tq+(1-t)p)$ hold. Specifically,*

$$
\begin{aligned}
(3.9) \qquad & |D_f(p, tq+(1-t)p) - f(1)| \\
\leq\ & t\|f''\|_\infty \left\{ \left(\frac{r+R}{2}-1\right) D_v(p,q) + \frac{t}{2}\tilde{D}(p,q) \right. \\
& \left. + (1-r)^2 \int_\chi |p(x)-q(x)|\, \Theta(x,t)\, d\mu(x) \right\} \\
\leq\ & t\|f''\|_\infty \left\{ \left[\frac{r+R}{2} - 1 + \frac{(1-r)^2}{(1-t)R+1-r}\right] D_v(p,q) + \frac{t}{2}\tilde{D}(p,q) \right\} \\
\leq\ & t\|f''\|_\infty \left[\frac{r+R}{2} - 1 + \frac{(1-r)^2}{(1-t)R+1-r} + (1-r)\frac{t}{2}\right] D_v(p,q),
\end{aligned}
$$

where

$$\Theta(x,t) = \frac{p(x)}{(R+t-r)\,p(x) - q(x)}$$

and $D_v(p,q)$ is the variational distance defined by (3.1) with $m=1$ and

$$\tilde{D}(p,q) := \int_\chi p(x)\left|1-\frac{q(x)}{p(x)}\right|\left(1-\frac{q(x)}{p(x)}\right) d\mu(x). \tag{3.10}$$

Proof. Substitution of $tq(x)+(1-t)p(x)$ for $q(x)$ in (3.8) gives on using the fact that $p(\cdot)$, $q(\cdot)$ are densities over the measure space μ on χ then

$$
\begin{aligned}
(3.11) \qquad & |D_f(p, tq+(1-t)p) - f(1)| \\
\leq\ & \|f''\|_\infty \int_\chi \left| \frac{tq(x)-(1-t)p(x)}{p(x)} - 1 \right| \\
& \times p(x)\, C\left(r, 1, \frac{tq(x)-(1-t)p(x)}{p(x)}, R\right) d\mu(x) \\
=\ & \|f''\|_\infty \int_\chi \frac{t}{p(x)} |p(x)-q(x)|\, H(p,q,x,t)\, d\mu(x),
\end{aligned}
$$

where

$$H(p,q,x,t) := p(x)\, C\left(r, 1, \frac{tq(x)-(1-t)p(x)}{p(x)}, R\right),$$

which, upon using (2.4) gives

$$
\begin{aligned}
& H(p,q,x,t) \\
= \; & \frac{1}{2}\frac{\left[(R+t-1)\,p(x)-tq(x)\right]^2(1-r)^2p^2(x)}{(R+t-1)\,p(x)-tq(x)+(1-r)\,p(x)} \\
= \; & \frac{(R+t-r)\,p(x)-tq(x)}{2}-(1-r)\,p(x)\left[\frac{(R+t-1)\,p(x)-tq(x)}{(R+t-r)\,p(x)-tq(x)}\right] \\
= \; & \frac{(R+t-r)\,p(x)-tq(x)}{2}-(1-r)\,p(x)+\frac{(1-r)^2p^2(x)}{(R+t-r)\,p(x)-tq(x)} \\
= \; & \left[\frac{r+R}{2}-1\right]p(x)+\frac{t}{2}\left(p(x)-q(x)\right)+\frac{(1-r)^2p^2(x)}{(R+t-r)\,p(x)-tq(x)}.
\end{aligned}
$$

Substitution of $H(p,q,x,t)$ into (3.11) gives the first inequality in (3.9). Now, for the second, we note that

$$
\frac{1}{(R+t-r)-tr}<\Theta(x,t)<\frac{1}{(1-t)\,R+t-r}
$$

since $r\leq\frac{q(x)}{p(x)}\leq R$, $x\in\chi$ and so utilising this upper bound in the first inequality of (3.9) produces the second.

The final inequality results on noting that

$$
1-R\leq 1-\frac{q(x)}{p(x)}\leq 1-r. \tag{3.12}
$$

The theorem is now completely proved. ∎

Remark 4. *Taking $t=1$ in the first and second results in (3.9) reproduce those presented in (3.4). The coarsest bounds for the two results are different. Further, taking $t=\frac{1}{2}$ in (3.9) gives*

$$
\begin{aligned}
& \left|D_f\left(p,\frac{q+p}{2}\right)-f(1)\right| \\
\leq \; & \frac{\|f''\|_\infty}{2}\left\{\left(\frac{r+R}{2}-1\right)D_v(p,q)\right. \\
& +\frac{1}{4}\int_\chi p(x)\left|1-\frac{q(x)}{p(x)}\right|\left(1-\frac{q(x)}{p(x)}\right)d\mu(x) \\
& \left.+(1-r)^2\int_\chi |p(x)-q(x)|\,\Theta\left(x,\frac{1}{2}\right)d\mu(x)\right\} \\
\leq \; & \frac{\|f''\|_\infty}{2}\left\{\left[\frac{r+R}{2}-1+\frac{(1-r)^2}{\frac{R}{2}+1-r}\right]D_v(p,q)\right. \\
& \left.+\frac{1}{4}\int_\chi p(x)\left|1-\frac{q(x)}{p(x)}\right|\left(1-\frac{q(x)}{p(x)}\right)d\mu(x)\right\} \\
\leq \; & \frac{\|f''\|_\infty}{2}\left\{\frac{r+R}{2}-1+(1-r)\left[\frac{1}{4}+\frac{2(1-r)}{R+2(1-r)}\right]\right\}.
\end{aligned}
$$

4. The Connection Between Integral Mean Divergence and the $f-$Divergence

It may have been noticed that there is a duality between $D_{\mathfrak{M}(g)}(p,q)$ and $D_f(p,q)$, the Integral Mean divergence and $f-$divergence examined in the previous sections.

We note that if

$$f(z) := \mathfrak{M}(g)(z), \tag{4.1}$$

where $\mathfrak{M}(g)(z)$ is as defined in (2.6) then

$$D_f(p,q) = D_{\mathfrak{M}(g)}(p,q) \tag{4.2}$$

and so the integral mean divergence is a particular instance of the $f-$divergence.

Contrarily, if we take

$$g(t) = f(t) + (t-1) f'(t) \tag{4.3}$$

then

$$\int_1^z g(t)\,dt = \int_1^z f(t)\,dt + \int_1^z (t-1) f'(t)\,dt$$

and so

$$\frac{1}{z-1}\int_1^z g(t)\,dt = f(z). \tag{4.4}$$

Further, allowing $z \to 1$ in (4.4) gives $g(0) = f(0)$.

Taking $z = \frac{q(x)}{p(x)}$ in (4.4) and multiplication by $p(x)$ over χ with respect to the measure μ then gives

$$D_{\mathfrak{M}(g)}(p,q) = D_f(p,q).$$

Thus the $f-$divergence is a particular case of the integral mean divergence with the relation (4.3) holding.

The above relationships hold for f and g replace by $f^{\dagger}$ and $g^{\dagger}$ where

$$h^{\dagger}(u) = h(u) + c(u-1), \quad c \in \mathbb{R}$$

since both $D_{\mathfrak{M}(g)}(p,q)$ and $D_f(p,q)$ satisfy the invariance property (p.3) as defined in the introduction.

The following lemma was proved in Qi et al. [12] which is useful for our discussion of the integral mean divergence.

Lemma 1. *Assume that if $f''(t)$ exists on $\mathbb{R}$ then if $f(t)$ is an increasing (convex) function on $\mathbb{R}$ then the arithmetic mean of $f(t)$*

$$\phi(r,s) = \begin{cases} \dfrac{1}{s-r}\displaystyle\int_r^s f(t)\,dt, & r \neq s, \\ \\ f(r), & r = s. \end{cases} \tag{4.5}$$

is also increasing (convex) with both r and s.

We note that the arithmetic mean (4.5) is equivalent to (2.2).

The above lemma implies that if $f''(t)$ exists on $\mathbb{R}$ then for convex $f(\cdot)$, $\mathcal{M}(f;r,s) \equiv \phi(r,s)$ is also convex in each direction. That is, if $f''(t)$ exists on $\mathbb{R}$ and $f(\cdot)$ is convex, then both $D_f(p,q)$ and $D_{\mathfrak{M}(f)}(p,q)$ are jointly convex in p and q ([4]).

5. Results for Functions of Bounded Variation

The following less restrictive result allowing a larger class of functions was proved in Cerone and Dragomir [11].

Theorem 4. *Let $g:[a,b]\to\mathbb{R}$ be of bounded variation on $[a,b]$ then for $a\le c<d\le b$*

$$\left|\frac{1}{b-a}\int_a^b g(t)\,dt-\frac{1}{d-c}\int_c^d g(t)\,dt\right|\le B(a,c,d,b)\,\frac{\bigvee_a^b(g)}{b-a} \tag{5.1}$$

where

$$B(a,c,d,b)=\frac{1}{2}\left[b-d+c-a+|b-d-(c-a)|\right] \tag{5.2}$$

and $\bigvee_a^b(g)$ represents the total variation of $g(t)$ over $[a,b]$.

The constant $\frac{1}{2}$ is best possible.

It should be noted that the requirement for $c<d$ was imposed for definiteness. All that is needed is that $c,d\in[a,b]$.

The following theorem holds for the integral mean divergence, $D_{\mathfrak{M}(g)}(p,q)$ defined by (2.5) – (2.6) where g is of bounded variation.

Theorem 5. *Let $g:\mathbb{R}\to\mathbb{R}$ be of bounded variation on any $[a,b]\subset\mathbb{R}$. If $p,q\in\Omega$ then with $0\le r\le 1\le R$, $r\le\frac{q(x)}{p(x)}\le R<\infty$ for $x\in\chi$ we have*

$$\begin{aligned}
&\left|D_{\mathfrak{M}(g)}(p,q)-\mathcal{M}(g;r,R)\right| \\
\le\ &\left[\frac{R-r}{2}+\int_\chi\left|\frac{R+r}{2}p(x)+\frac{p(x)+q(x)}{2}\right|d\mu(x)\right]\frac{\bigvee_r^R(g)}{R-r} \\
\le\ &\frac{R+1}{R-r}\bigvee_r^R(g),
\end{aligned} \tag{5.3}$$

where $\mathfrak{M}(g)$ is as defined by (2.6).

Proof. Let $a=r$, $b=R$ and $c=1$, $d=z$ then from (5.1) we have

$$|\mathcal{M}(g;1,z)-\mathcal{M}(g;r,R)|\le B(r,1,z,R)\,\frac{\bigvee_r^R(g)}{R-r}$$

which, upon taking $z=\frac{q(x)}{p(x)}$ for $x\in\chi$ gives

$$\left|\mathcal{M}\left(g;1,\frac{q(x)}{p(x)}\right)-\mathcal{M}(g;r,R)\right|\le B\left(r,1,\frac{q(x)}{p(x)},R\right)\frac{\bigvee_r^R(g)}{R-r}, \tag{5.4}$$

where $B(a,c,d,b)$ is given by (5.2).

Multiplication of (5.4) by $p(x)$ for $x\in\chi$ and integration with respect to the measure μ produces the first of the inequalities in (5.3). The second is obtained by using the fact that $|A-B|\le|A|+|B|$ and that both $p,q\in\Omega$ defined by (1.1). ∎

Corollary 3. *Let the conditions of Theorem 4 hold, then for $t\in[0,1]$, the parameterised integral divergence measure $D_{\mathfrak{M}(g)}(p,tq+(1-t)p)$ satisfies the following* *inequalities. Namely,*

$$
\begin{aligned}
(5.5)\qquad & \left|D_{\mathfrak{M}(g)}\left(p, tq+(1-t)\,p\right)-\mathcal{M}\left(g;r,R\right)\right| \\
\leq\; & \left[\frac{R-r}{2}+\int_{\chi}\left|\left(\frac{R+r}{2}-1\right)p\left(x\right)-\frac{t}{2}\left(q\left(x\right)-p\left(x\right)\right)\right|d\mu\left(x\right)\right]\frac{\bigvee_r^R\left(g\right)}{R-r} \\
\leq\; & \left[\max\left\{1-r,R-1\right\}+\frac{t}{2}D_v\left(p,q\right)\right]\frac{\bigvee_r^R\left(g\right)}{R-r},
\end{aligned}
$$

where $D_v(p,q)$ is the variational distance defined by (3.6).

Proof. Substitution of $tq(x)+(1-t)\,p(x)$ for $q(x)$ in the first inequality of (5.3) produces the first inequality in (5.5).

The second is obtained from the fact that $|A-B|\leq|A|+|B|$ so that

$$
\begin{aligned}
& \int_{\chi}\left|\left(\frac{R+r}{2}-1\right)p\left(x\right)-\frac{t}{2}\left(q\left(x\right)-p\left(x\right)\right)\right|d\mu\left(x\right) \\
\leq\; & \left|\frac{R+r}{2}-1\right|\int_{\chi}p\left(x\right)d\mu\left(x\right)+\frac{t}{2}\int_{\chi}\left|p\left(x\right)-q\left(x\right)\right|d\mu\left(x\right).
\end{aligned}
$$

Now, since $\int_{\chi}p(x)\,d\mu(x)=1$ and $\frac{R-r}{2}+\left|\frac{R+r}{2}-1\right|=\max\{1-r,R-1\}$ the second inequality as stated in (5.5) results. ∎

Remark 5. *We note that for $R-1>1-r$ then the second (or third) inequality in (5.5) provides a tighter bound than the second in (5.3).*

The following theorem provides approximations for the $f-$divergence.

Theorem 6. *Let $g:\mathbb{R}\to\mathbb{R}$ be of bounded variation on $[a,b]\subset\mathbb{R}$ then if $p,q\in\Omega$ and $0\leq r\leq1\leq R$, $r\leq\frac{q(x)}{p(x)}\leq R<\infty$ for $x\in\chi$ we have, for $g(t)=f'(t)$,*

$$
\begin{aligned}
(5.6)\qquad & \left|D_f\left(p,q\right)-f\left(1\right)\right| \\
\leq\; & \left\{\frac{R-r}{2}D_v\left(p,q\right)+\frac{1}{2}\tilde{D}\left(p,q\right)\right. \\
& \left.+\frac{1}{2}\int_{\chi}p\left(x\right)\left|1-\frac{q\left(x\right)}{p\left(x\right)}\right|\left|R-\left(1-r\right)-\frac{q\left(x\right)}{p\left(x\right)}\right|d\mu\left(x\right)\right\}\frac{\bigvee_r^R\left(f'\right)}{R-r} \\
\leq\; & \left\{\left[\frac{R-r}{2}+\frac{1}{2}\max\left\{R-1,1-r\right\}\right]D_v\left(p,q\right)-\frac{1}{2}\tilde{D}\left(p,q\right)\right\}\frac{\bigvee_r^R\left(f'\right)}{R-r} \\
\leq\; & \left[\frac{3}{4}\left(R-r\right)+\frac{1-r}{2}+\frac{1}{2}\left|\frac{r+R}{2}-1\right|\right]\frac{\bigvee_r^R\left(f'\right)}{R-r},
\end{aligned}
$$

where $D_v(p,q)$ is as given by (3.1) with $m=1$ and $\tilde{D}(p,q)$ is defined by (3.10).

Proof. The proof follows that of Theorem 4. Taking $a=r$, $b=R$ and $c=1$, $d=z$ in (5.1) gives with $g(t)=f'(t)$

$$
(5.7)\qquad \left|f\left(z\right)-f\left(1\right)-\left(z-1\right)\left(\frac{f\left(R\right)-f\left(r\right)}{R-r}\right)\right|\leq\left|z-1\right|B\left(r,1,z,R\right)\frac{\bigvee_r^R\left(f'\right)}{R-r},
$$

with $B(a,c,d,b)$ being defined by (5.2).

Taking $z = \frac{q(x)}{p(x)}$ for $x \in \chi$ in (5.7), multiplication by $p(x)$ and integration over χ with respect to the measure μ gives

$$\begin{aligned} &|D_f(p,q) - f(1)| \\ &\leq \frac{\bigvee_r^R (f')}{R-r} \int_\chi \left| \frac{q(x)}{p(x)} - 1 \right| p(x) B\left(r, 1, \frac{q(x)}{p(x)}, R\right) d\mu(x). \end{aligned} \tag{5.8}$$

Now, from (5.2)

$$\begin{aligned} p(x) B\left(r, 1, \frac{q(x)}{p(x)}, R\right) &= \frac{R+r}{2} p(x) - \frac{q(x) - p(x)}{2} \\ &\quad + \frac{1}{2} |(R - (1-r)) p(x) - q(x)| \end{aligned}$$

and so substitution into (5.8) gives the first inequality in (5.6).

For the second inequality, we observe that since $r \leq \frac{q(x)}{p(x)} \leq R$ then

$$-(1-r) \leq R - (1-r) - \frac{q(x)}{p(x)} \leq R - 1$$

giving

$$\left| R - (1-r) - \frac{q(x)}{p(x)} \right| \leq \max\{R-1, 1-r\}. \tag{5.9}$$

Using (5.9) in the first inequality of (5.6) gives the second.

The final inequality results from (3.12) so that

$$1 - \frac{q(x)}{p(x)} \leq 1 - r$$

and so

$$\begin{aligned} \frac{1}{2} \tilde{D}(p,q) &= \frac{1}{2} \int_\chi p(x) \left| 1 - \frac{q(x)}{p(x)} \right| \left(1 - \frac{q(x)}{p(x)}\right) d\mu(x) \\ &\leq \frac{1-r}{2} D_v(p,q) \end{aligned}$$

producing the third inequality in (5.6). ∎

Corollary 4. *Let the conditions of Theorem 6 hold, then for $t \in [0,1]$ the parametrised version of the divergence measure introduced by Csiszár, $D_f(p, tq + (1-t)p)$ satisfies the following results. That is,*

$$\begin{aligned} &|D_f(p, tq + (1-t)p) - f(1)| \\ &\leq \frac{t}{2} \left\{ (R-r) D_v(p,q) + t \int_\chi p(x) \left| 1 - \frac{q(x)}{p(x)} \right| \left(1 - \frac{q(x)}{p(x)}\right) d\mu(x) \right. \\ &\quad \left. + \int_\chi p(x) \left| 1 - \frac{q(x)}{p(x)} \right| \left| r + R + t - 2 - t\left(\frac{q(x)}{p(x)}\right) \right| d\mu(x) \right\} \frac{\bigvee_r^R (f')}{R-r} \\ &\leq \frac{t}{2} \left[(R-r) + t(1-r) + \max\{A,B\} \right] D_v(p,q) \frac{\bigvee_r^R (f')}{R-r}, \end{aligned} \tag{5.10}$$

where $\max\{A,B\} = \max\{|(1-t)(1-\lambda) - \lambda|, |1 - \lambda - \lambda(1-t)|\}$ *with* $\lambda = \frac{1-r}{R-r}$.

Proof. We commence from (5.8) with $q(x)$ being replaced by $tq(x)+(1-t)p(x)$ so that

$$|D_f(p,tq+(1-t)p)-f(1)| \le \frac{\bigvee_r^R(f')}{R-r}t\int_\chi |p(x)-q(x)|\,B\left(r,1,\frac{tq(x)+(1-t)p(x)}{p(x)},R\right)d\mu(x). \tag{5.11}$$

Now, from (5.2),

$$\begin{aligned} & p(x)\,B\left(r,1,\frac{tq(x)+(1-t)p(x)}{p(x)},R\right) \\ = & \frac{1}{2}\left\{(R+t-r)\,p(x)-tq(x)+|(r+R+t-2)\,p(x)-tq(x)|\right\} \end{aligned}$$

and so upon substitution into (5.11) produces the first inequality.

Now, from (3.12) we have

$$1-R\le 1-\frac{q(x)}{p(x)}\le 1-r \tag{5.12}$$

and so

$$-t(R-1)\le t\left(1-\frac{q(x)}{p(x)}\right)\le t(1-r)$$

and

$$\begin{aligned}(1-t)(R-1)-(1-r) &\le R-1-(1-r)+t\left(1-\frac{q(x)}{p(x)}\right)\\ &\le (R-1)-(1-t)(1-r).\end{aligned}$$

Let $\lambda=\frac{1-r}{R-r}$ and so $\frac{R-1}{R-r}=1-\lambda$ giving

$$\begin{aligned}(1-t)(1-\lambda)-\lambda &\le 1-2\lambda+t\left(1-\frac{q(x)}{p(x)}\right)\\ &\le 1-\lambda-\lambda(1-t).\end{aligned}$$

That is,

$$\begin{aligned} & \left|1-2\lambda+t\left(1-\frac{q(x)}{p(x)}\right)\right| \\ \le & \max\left\{|(1-t)(1-\lambda)-\lambda|,|1-\lambda-\lambda(1-t)|\right\}. \end{aligned} \tag{5.13}$$

Using (5.11) and (5.13) in the first inequality of (5.10) produces the second. ∎

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] I. CSISZÁR and J. KÖRNER, *Information Theory: Coding Theorem for Discrete Memoryless Systems,* Academic Press, New York, 1981.

[5] S.S. DRAGOMIR, On the Ostrowski's integral inequality for mappings of bounded variation and applications, *RGMIA Res. Rep. Coll.,* **2**(1) (1999), 73-80.

[6] S.S. DRAGOMIR, V. GLUŠČEVIĆ and C.E.M. PEARCE, Csiszár $f-$divergence, Ostrowski's inequality and mutual information, to appear in *J. Nonlin. Anal. Ser. A.*

[7] S.S. DRAGOMIR, V. GLUŠČEVIĆ and C.E.M. PEARCE, Approximations for Csiszár $f-$divergence via midpoint inequalities.

[8] S.S. DRAGOMIR, V. GLUŠČEVIĆ and C.E.M. PEARCE, The approximations for Csiszár $f-$divergence for absolutely continuous mappings.

[9] S.S. DRAGOMIR, V. GLUŠČEVIĆ and C.E.M. PEARCE, The approximations for Csiszár $f-$divergence for mappings of bounded variation.

[10] N.S. BARNETT, P. CERONE, S.S. DRAGOMIR and A.M. FINK, Comparing two integral means for absolutely continuous mappings whose derivatives are in $L_\infty[a,b]$ and applications, *Comp. and Math. with Applics.,* (accepted).

[11] P. CERONE and S.S. DRAGOMIR, Differences between means with bounds from a Riemann-Stieltjes integral, *Comp. and Math. with Applics.,* (accepted).

[12] F. QI, S.-L. XU and L. DEBNATH, A new proof of monotonicity for extended mean values, *Intern. J. Math. Sci.,* **22**(2) (1999), 415-420.

[13] H. SHIOYA and T. DA-TE, A generalisation of Lin divergence and the derivation of a new information divergence measure, *Electronics and Communications in Japan,* **78**(7) (1995), 34-40.

[14] H. SHIOYA, H. NAGAOKA and T. DA-TE, A new inequality for $f-$divergence, its maximum, and application to learning problems, *Trans.(A), I.E.I.C.E.,* Japan, J73-74(4) (1994), 720-726.

[15] J. LIN, Divergence measures based on the Shannon entropy, *IEEE Trans., IT,* **37**(1) (1991), 145-151.

[16] J. LIN and S.K.M. WONG, A new directed divergence measure and its characterisation, *Int. J. General Systems,* **17** (1990), 73-81.

[17] J.N. KAPUR, A comparative assessment of various measures of directed divergence, *Advances in Management Studies,* **3** (1984), 1-16.

[18] I.J. TANEJA, *Generalised Information Measures and Their Applications*, (http://www.mtm.ufsc.br/~taneja/bhtml/bhtml.html)

School of Communications and Informatics, Victoria University of Technology, PO Box 14428, Melbourne City MC, 8001, Victoria, Australia.

E-mail address: pc@matilda.vu.edu.au
URL: http://rgmia.vu.edu.au/cerone

E-mail address: sever@matilda.vu.edu.au
URL: http://rgmia.vu.edu.au/SSDragomirWeb.html