New Approximations for f-Divergence via Trapezoid and Midpoint Inequalities

NEW APPROXIMATIONS FOR $f-$DIVERGENCE VIA TRAPEZOID AND MIDPOINT INEQUALITIES

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

Abstract. Using sharp inequalities of trapezoid and midpoint type in terms of the infinum and supremum of the derivative, some new and better approximation of $f-$divergence are given. Application for some particular instances are also mentioned.

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 $\{a_i | i = 1, ..., n\}$, 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 [1]) are defined respectively as

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

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

Csizar [3] - [4] has introduced a versatile functional from which subsumes a number of the more popular choices of divergence measures, including those mentioned above. For a convex function $f : [0, \infty) \to R$, the f*-divergence* between p and q is defined by (see also [5])

$$I_f(p, q) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i} \right) . \tag{1.3}$$

It is convenient to invoke as a benchmark the chi-squared discrepancy measure

$$D_{\chi^2}(p, q) := \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} = \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1 \tag{1.4}$$

which arises from (1.3) as the particular case $f(x) = (x - 1)^2$.

Date: December 6, 2002.

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

Key words and phrases. $f-$divergence, midpoint inequalities, trapezoid inequalities, divergence measures.

Most common choices of f, like the above, satisfy $f(1)=0$, so that $I_f(q,p)=0$. Convexity then ensures that $I_f(q,p)$ is nonnegative. However, as noted in [2], some additional flexibility for applications can be achieved by not insisting on convexity.

For other properties of f-divergence and applications, see [6] and the references therin.

By the use of mid-point inequality, the following result may be stated (see also [7])

Theorem 1. *Assume that $p=(p_1,\dots,p_n)$, $q=(q_1,\dots,q_n)$ are probability distributions satisfying the assumptions*

$$0 \le r \le \frac{p_i}{q_i} \le R \le \infty \quad (\text{where } r \le 1 \le R \text{) for each } i \in \{1,...,n\}. \tag{1.5}$$

If $f:[0,\infty)\to R$ is so that is locally absolutely continuous in $[r,R)$ and $f^{''}\in L_\infty[r,R)$, then

$$\left|I_f(p,q)-f(1)-I_{f_b}(p,q)\right| \le \frac{1}{4}\left\|f^{''}\right\|_{[r,R),\infty} D_{\chi^2}(p,q) \tag{1.6}$$

where $f_b(x)=(x-1)f'\left(\frac{x+1}{2}\right)$, $x\in[r,R)$.

Using Iyengar inequality that provides a refinement of the trapezoid inequality, the following result also holds [8]

Theorem 2. *With the assumptions in Theorem 1 one has*

$$\begin{aligned} &\left|I_f(p,q)-f(1)-\frac{1}{2}I_{f_\#}(p,q)\right| \\ \le\ & \frac{1}{4}\left\|f^{''}\right\|_{[r,R),\infty} D_{\chi^2}(p,q)-\frac{1}{4\left\|f^{''}\right\|_{[r,R),\infty}} I_{f_0}(p,q) \\ \le\ & \frac{1}{4}\left\|f^{''}\right\|_{[r,R),\infty} D_{\chi^2}(p,q) \end{aligned} \tag{1.7}$$

where $f_\#(x)=(x-1)f'(x)$ and $f_0(x)=\left|f'(x)-f'(1)\right|^2$, $x\in[r,R)$.

In this paper similar bounds are provided when information about $\gamma=\inf_{t\in[r,R)} f^{''}(t)$ and $\Gamma=\sup_{t\in[r,R)} f^{''}(t)$ are assumed to be known.

Applications for particular instances of $f-$divergences are also pointed out.

2. Some General Bounds for $f-$Divergence

The following analytic inequality is useful in the following. It has been obtained in [9] with a different proof than provided here for the sake of completeness.

Lemma 1. *Let $\varphi:[a,b]\to R$ be an absolutely continuous function on $[a,b]$ with the property that there exists the constants $m,M\in R$ with*

$$m \le \varphi'(t) \le M \quad \text{for all } t\in[a,b]. \tag{2.1}$$

Then we have the inequality

$$\left|\frac{\varphi(a)+\varphi(b)}{2}-\frac{1}{b-a}\int_a^b \varphi(t)\,d(t)\right| \le \frac{1}{8}(M-m)(b-a). \tag{2.2}$$

The constant $\frac{1}{8}$ is best possible in the sense that it can not be replaced by a smaller constant.

Proof. Start to the following identity that obviously holds integrating by parts

$$\frac{\varphi(a)+\varphi(b)}{2}-\frac{1}{b-a}\int_a^b\varphi(t)\,d(t)=\frac{1}{b-a}\int_a^b\left(t-\frac{a+b}{2}\right)\varphi'(t)dt. \tag{2.3}$$

Observe that

$$\frac{1}{b-a}\int_a^b\left(t-\frac{a+b}{2}\right)\varphi'(t)dt=\frac{1}{b-a}\int_a^b\left(t-\frac{a+b}{2}\right)\left(\varphi'(t)-\frac{m+M}{2}\right)dt$$

and since

$$\left|\varphi'(t)-\frac{m+M}{2}\right|\le\frac{M-m}{2}\quad\text{for all}\quad t\in[a,b]$$

we deduce

$$\begin{aligned}&\frac{1}{b-a}\left|\int_a^b\left(t-\frac{a+b}{2}\right)\left(\varphi'(t)-\frac{m+M}{2}\right)dt\right| \\ \le\;&\frac{1}{b-a}\frac{M-m}{2}\int_a^b\left|t-\frac{a+b}{2}\right|dt \\ =\;&\frac{M-m}{8}(b-a).\end{aligned} \tag{2.4}$$

Since the case of equality in (2.2) is realised for the absolutely continuous function $\varphi_0:[a,b]\to m,\ \ \varphi_0(t)=k\left|t-\frac{a+b}{2}\right|,\ k>0$, the sharpness of the constant easily follows, and we omit the details. ∎

For a differentiable function $f:[0,\infty)\to R$, consider the associated function $f_{\#}:(0,\infty)\to R$ given by

$$f_{\#}(u):=(u-1)f'(u),\qquad u\in(0,\infty). \tag{2.5}$$

The following result holds.

Theorem 3. *Assume that $p=(p_1,\dots,p_n)$, $q=(q_1,\dots,q_n)$ are probability distributions satisfying the assumption*

$$0\le r\le\frac{p_i}{q_i}\le R\le\infty\quad(\textit{where } r\le1\le R\ )\textit{ for each } i\in\{1,...,n\}\,. \tag{2.6}$$

If $f:[0,\infty)\to R$ is so that f' is locally absolutely continuous on $[\gamma,R)$ and there exists the real numbers γ,Γ so that

$$\gamma\le f''(t)\le\Gamma\quad\textit{for all } t\in(r,R); \tag{2.7}$$

then one has the inequality

$$\left|I_f(p,q)-f(1)-\frac{1}{2}I_{f_{\#}}(p,q)\right|\le\frac{1}{8}(\Gamma-\gamma)D_{\chi^2}(p,q). \tag{2.8}$$

Proof. Applying the inequality (2.2) for $\varphi(t)=f'(t)$, $b=x\in(r,R)$, $a=1$, $M=\Gamma$ and $m=\gamma$, we deduce

$$\left|f(x)-f(1)-\frac{1}{2}(x-1)\left(f'(1)+f'(x)\right)\right|\le\frac{1}{8}(\Gamma-\gamma)(x-1)^2 \tag{2.9}$$

for any $x\in(r,R)$ (and if $\gamma=0$ and $R=\infty$, for any $x\in(0,\infty)$).

Choose in (2.9) $r = \frac{p_i}{q_i}$ $(i = 1, ..., n)$ and multiply by $q_i \geq 0$ $(i = 1, ..., n)$ to get

$$\left| q_i f\left(\frac{p_i}{q_i}\right) - f(1) q_i - \frac{1}{2}\left(\frac{p_i}{q_i} - 1\right) f'(1) q_i - \frac{1}{2}\left(\frac{p_i}{q_i} - 1\right) f'\left(\frac{p_i}{q_i}\right) q_i \right| \leq \frac{1}{8}(\Gamma - \gamma) q_i \left(\frac{p_i}{q_i} - 1\right)^2 \tag{2.10}$$

for any $i \in \{1, ..., n\}$. If we sum in (2.10) over i from 1 to n and take into account that $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$, then by the generalized triangle inequality we deduce the desired result (2.8). ∎

Remark 1. *The inequality (2.8) is an improvement of (1.6) since* $0 \leq \Gamma - \gamma \leq 2\left\| f'' \right\|_{[r,R),\infty}$.

To establish our second result, we need the following inequality obtained in [9] for which we give here a simple direct proof.

Lemma 2. *Assume that φ is as in Lemma 1. Then one has the inequality*

$$\left| \varphi\left(\frac{a+b}{2}\right) - \frac{1}{b-a}\int_a^b \varphi(t)\, d(t) \right| \leq \frac{1}{8}(M - m)(b - a). \tag{2.11}$$

The constant $\frac{1}{8}$ is best possible in the sense mentioned in Lemma 1.

Proof. Start to the following identity that obviously holds integrating by parts

$$\varphi\left(\frac{a+b}{2}\right) - \frac{1}{b-a}\int_a^b \varphi(t)\, d(t) = \frac{1}{b-a}\int_a^b K(t)\, \varphi'(t) d(t) \tag{2.12}$$

where

$$K(t) = \begin{cases} t - a & \text{if} \quad t \in \left[a, \frac{a+b}{2}\right] \\ t - b & \text{if} \quad t \in \left[\frac{a+b}{2}, b\right] \end{cases}.$$

Since

$$\int_a^b K(t)\, d(t) = 0,$$

we observe that

$$\frac{1}{b-a}\int_a^b K(t)\varphi'(t)\, d(t) = \frac{1}{b-a}\int_a^b K(t)\left(\varphi'(t) - \frac{m+M}{2}\right) d(t)$$

and since

$$\left| \varphi'(t) - \frac{m+M}{2} \right| \leq \frac{M-m}{2} \quad \text{for all} \quad t \in [a, b],$$

we deduce

$$\begin{aligned} \frac{1}{b-a}\left| \int_a^b K(t)\left(\varphi'(t) - \frac{m+M}{2}\right) dt \right| & \\ \leq \frac{1}{b-a}\frac{M-m}{2}\int_a^b |K(t)|\, dt & \\ = \frac{1}{8}(M - m)(b - a). & \end{aligned} \tag{2.13}$$

Since the case of equality in (2.11) is realised for the absolutely continuous function $\varphi_0 : [a,b] \to R$, $\varphi_0(t) = k\left|t - \frac{a+b}{2}\right|$, $k > 0$, the sharpness of the constant is proved and we omit the details. ∎

For a differentiable function $f : [0,\infty) \to R$, consider now the associated function $f_b : (0,\infty) \to R$, given by

$$f_b(x) := (x-1)f'\left(\frac{x+1}{2}\right). \tag{2.14}$$

The following result holds.

Theorem 4. *Assume that p, q, f, γ and Γ are as in Theorem 2. Then one has the inequality*

$$\left|I_f(p,q) - f(1) - I_{f_b}(p,q)\right| \le \frac{1}{8}(\Gamma-\gamma)D_{\chi^2}(p,q). \tag{2.15}$$

Proof. Applying the inequality (2.11) for $\varphi(t) = f'(t)$, $b = x \in (r,R)$, $a = 1$, $M = \Gamma$ and $m = \gamma$, we deduce

$$\left|f(x) - f(1) - (x-1)f'\left(\frac{x+1}{2}\right)\right| \le \frac{1}{8}(\Gamma-\gamma)(x-1)^2. \tag{2.16}$$

for any $x \in (r,R)$ (and if $r = 0$ and $R = \infty$, for any $x \in (0,\infty)$).

Making use of the same argument utilized in the proof of Theorem 2, we deduce the desired result (2.15). ∎

Remark 2. *The inequality (2.15) provides a different bound then (1.2). The bound provided by (2.15) is better then the second bound in (1.7) since in general $0 \le \Gamma - \gamma \le 2\left\|f''\right\|_{[r,R),\infty}$.*

3. Applications

(1) The Kullback-Leibler divergence $D(p,q)$ is generated by the convex function $f(u) = u\ln u$, $u \in (0,\infty)$. Obviously

$$f_\#(u) = (u-1)\ln u + u - 1, \quad u \in (0,\infty).$$

We observe that

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

Observe also that $f''(u) = \frac{1}{u}$ and if $0 < r \le u \le R \le 0$, $i = 1,...,n$; then

$$\frac{1}{R} \le f''(u) \le \frac{1}{r}, \quad \text{for} \quad u \in [r,R].$$

Using the inequality (2.8) we deduce

$$\left|D(p,q) - \frac{1}{2}\left[D(p,q) + D(q,p)\right]\right| \le \frac{1}{8}\left(\frac{1}{r} - \frac{1}{R}\right)D_{\chi^2}(p,q)$$

giving the following inequality

$$|D(p,q) - D(q,p)| \le \frac{1}{4}\frac{R-r}{rR} D_{\chi^2}(p,q) \tag{3.1}$$

for any p, q probability distributions provided

$$0 < r \le \frac{p_i}{q_i} \le R < \infty, \quad \text{for each} \quad i \in \{1, ..., n\}. \tag{3.2}$$

Now observe that

$$f_b(u) := (u-1)\ln\left(\frac{1+u}{2}\right) + u - 1, \quad u \in (0,\infty).$$

We observe that

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

Utilizing (2.15) we can conclude that

$$|D(p,q) - K(q,p)| \le \frac{1}{8}\frac{R-r}{rR} D_{\chi^2}(p,q) \tag{3.3}$$

provided p, q satisfy (3.2).

(2) Consider the convex function $f : (0,\infty) \to R,\ f(x) = -\ln x$. Then

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

Observe also that

$$f_{\#}(u) = \frac{1-u}{u}.$$

We have

$$I_{f_{\#}}(p,q) = \sum_{i=1}^n q_i \left(\frac{1-\frac{p_i}{q_i}}{\frac{p_i}{q_i}}\right) = \sum_{i=1}^n \frac{q_i^2}{p_i} - 1 = D_{\chi^2}(p,q).$$

Since $f''(u) = \frac{1}{u^2}$ and for $0 < r \le u \le R < \infty$ one has $\frac{1}{R^2} \le f''(u) \le \frac{1}{r^2}$, then by inequality (2.2) we deduce

$$\left|D(p,q) - \frac{1}{2}D_{\chi^2}(p,q)\right| \le \frac{1}{8}\frac{R^2 - r^2}{r^2R^2} D_{\chi^2}(p,q) \tag{3.4}$$

provided p, q satisfy (3.2).

Now, observe that

$$f_b(u) = \frac{2(1-u)}{u+1}.$$

For this function we have

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

Using the inequality (2.15) we deduce

$$|D(p,q) - L(p,q)| \le \frac{1}{8}\frac{R^2 - r^2}{r^2R^2} D_{\chi^2}(p,q) \tag{3.5}$$

provided p, q satisfy (3.2).

(3) Consider the function $f(u) = \sqrt{1+u^2} - \frac{1+u}{\sqrt{2}}$. Then $f'(u) = \frac{u}{\sqrt{1+u^2}} - \frac{\sqrt{2}}{2}$ and $f''(u) = \frac{1}{(1+u^2)\sqrt{1+u^2}}$.

The f-divergence introduced by this function is the "perimeter divergence" and has been considered in 1982 by F. Österreicher [10] . We obviously have

$$P(p,q) = \sum_{i=1}^{n} q_i \left[\sqrt{1 + \left(\frac{p_i}{q_i}\right)^2} - \frac{1 + \frac{p_i}{q_i}}{\sqrt{2}} \right] = \sum_{i=1}^{n} \sqrt{p_i^2 + q_i^2} - \sqrt{2}. \tag{3.6}$$

Observe that

$$f_{\#}(u) = (u-1)f'(u) = \frac{u(u-1)}{\sqrt{1+u^2}} - \frac{\sqrt{2}}{2}(u-1)$$

and thus

$$\begin{aligned} I_{f\#}(p,q) &= \sum_{i=1}^{n} q_i \frac{\frac{p_i}{q_i}\left(\frac{p_i}{q_i} - 1\right)}{\sqrt{1 + \left(\frac{p_i}{q_i}\right)^2}} = \sum_{i=1}^{n} \frac{p_i(p_i - q_i)}{\sqrt{q_i^2 + p_i^2}} \\ &= \sum_{i=1}^{n} \frac{p_i^2 + q_i^2 - p_i q_i - q_i^2}{\sqrt{q_i^2 + p_i^2}} = \sum_{i=1}^{n} \sqrt{q_i^2 + p_i^2} - \sum_{i=1}^{n} \frac{q_i(p_i + q_i)}{\sqrt{q_i^2 + p_i^2}}. \end{aligned} \tag{3.7}$$

Define

$$\begin{aligned} S(p,q) &= \sqrt{2} - \sum_{i=1}^{n} \frac{q_i(p_i + q_i)}{\sqrt{q_i^2 + p_i^2}} \\ &= \sum_{i=1}^{n} q_i \left[\frac{\sqrt{2}\sqrt{p_i^2 + q_i^2} - (p_i + q_i)}{\sqrt{q_i^2 + p_i^2}} \right] \geq 0. \end{aligned} \tag{3.8}$$

Then, by (3.2), we have

$$I_{f\#}(p,q) = P(p,q) + S(p,q).$$

We also observe that $0 \leq f''(u) \leq 1$ for any $u \in [0, \infty)$, and thus by (2.8) one has the inequality

$$|P(p,q) - S(p,q)| \leq \frac{1}{4} D_{\chi^2}(p,q) \tag{3.9}$$

for any p, q probability distributions.

References

[1] S. KULLBACK and R.A. LEIBLER, On information and sufficiency, *Ann. Math. Stat.*, **22** (1951), 79-86.

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

[3] I. CSISZÁR, Information-type measures of difference of probability distributions and indirect observations, *Studia Math. Hungarica,* **2** (1967), 299-318.

[4] I. CSISZÁR, On topological properties of $f-$divergences, *Studia Math. Hungarica,* **2** (1967), 329-339.

[5] S.M. ALI and S.D. SILVEY, A general class of coefficients of divergence of one distribution from another, *J. Roy. Statist. Soc. Sec B,* **28** (1966), 131-142.

[6] I. VAJDA, *Theory of Statistical Inference and Information,* Dordrecht-Boston, Kluwer Academic Publishers, 1989.

[7] S.S. DRAGOMIR, V. GLUSCEVIC and C.E. M. PEARCE, Approximations for the Csiszàr f-divergence via midpoint inequalities, in *Inequality Theory and Applications* - **Volume 1**, Y.J. Cho, J.K. Kim and S.S. Dragomir (Eds.), Nova Science Publishers, Inc., Huntington, New York, 2001, pp. 139-154.

[8] S.S. DRAGOMIR, V. GLUSCEVIC and C.E. M. PEARCE, The approximation of Csiszàr f-divergence for absolutely continuous mappings, in *Inequality Theory and Applications* - **Volume 1**, Y.J. Cho, J.K. Kim and S.S. Dragomir (Eds.), Nova Science Publishers, Inc., Huntington, New York, 2001, pp. 155-170.

[9] S.S. DRAGOMIR, The Median Principle for Inequalities and Applications, Preprint, *RGMIA Res. Rep. Coll.* , **5**(2002), Supplement, Article 17, [ON LINE `http://rgmia.vu.edu.au/v5(E).html`]

[10] F. ÖSTERREICHER, The construction of least favourable distributions is traceable to a minimal perimeter problem, *Studia Sci. Math. Hungar.*, **17**(1982), 341-351.

School of Communications and Informatics, Victoria University of Technology, PO Box 14428, Melbourne City MC, Victoria 8001, Australia.
E-mail address: `sever@matilda.vu.edu.au`
URL: `http://rgmia.vu.edu.au/SSDragomirWeb.html`

Royal Australian Airforce, ARDU, PO Box 1500, Salisbury SA 5108, Australia.
E-mail address: `vido@senet.com.au`

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