APPROXIMATING TWO MAPPINGS ASSOCIATED TO CSISZÁR $f-$DIVERGENCE VIA TAYLOR'S EXPANSION

N.S. BARNETT, P. CERONE, S.S. DRAGOMIR, AND A. SOFO

Abstract. Using the Taylor representation with integral remainder, we point out some approximations of two mappings generalising Csiszár $f-$divergence.

1. Introduction

One of the important issues in many applications of Probability Theory is finding an appropriate measure of *distance* (or *difference* or *discrimination*) between two probability distributions. A number of divergence measures for this purpose have been proposed and extensively studied by Jeffreys [1], Kullback and Leibler [2], Rényi [3], Havrda and Charvat [4], Kapur [5], Sharma and Mittal [6], Burbea and Rao [8], Rao [9], Lin [10], Csiszár [11], Ali and Silvey [13], Vajda [14], Shioya and Da-te [14] and others (see for example [5] and the references therein).

These measures have been applied in a variety of fields such as: anthropology [9], genetics [15], finance, economics, and political science [16], [17], [18], biology [19], the analysis of contingency tables [20], approximation of probability distributions [21], [22], signal processing [23], [24] and pattern recognition [25], [26]. A number of these measures of distance are specific cases of Csiszár f-divergence and so further exploration of this concept will have a flow on effect to other measures of distance and to areas in which they are applied.

Assume that a set Γ and the $\sigma-$finite measure μ are given. Consider the set of all probability densities on μ to be $\Omega := \left\{p | p : \Gamma \to \mathbb{R},\, p(x) \geq 0,\ \int_{\Gamma} p(x)\, d\mu(x) = 1\right\}$. The Kullback-Leibler divergence [2] is well known among the information divergences. It is defined as:

$$D_{KL}(p,q) := \int_{\Gamma} p(x) \log\left[\frac{p(x)}{q(x)}\right] d\mu(x), \quad p,q \in \Omega, \tag{1.1}$$

where log is to base 2.

In Information Theory and Statistics, various divergences are applied in addition to the Kullback-Leibler divergence. These are the: *variation distance* D_v, *Hellinger distance* D_H [2], χ^2-*divergence* D_{χ^2}, $\alpha-$*divergence* D_α, *Bhattacharyya distance* D_B [3], *Harmonic distance* D_{Ha}, *Jeffreys distance* D_J [1], *triangular discrimination* D_Δ [36], etc... They are defined as follows:

$$D_v(p,q) := \int_{\Gamma} |p(x) - q(x)|\, d\mu(x), \quad p,q \in \Omega; \tag{1.2}$$

Date: April 11, 2000.
1991 *Mathematics Subject Classification.* Primary 94Xxx; Secondary 26D15.
Key words and phrases. Csiszár $f-$divergegence, Taylor's formula.

$$D_H(p,q) := \int_\Gamma \left|\sqrt{p(x)} - \sqrt{q(x)}\right| d\mu(x), \quad p,q \in \Omega; \tag{1.3}$$

$$D_{\chi^2}(p,q) := \int_\Gamma p(x)\left[\left(\frac{q(x)}{p(x)}\right)^2 - 1\right] d\mu(x), \quad p,q \in \Omega; \tag{1.4}$$

$$D_\alpha(p,q) := \frac{4}{1-\alpha^2}\left[1 - \int_\Gamma [p(x)]^{\frac{1-\alpha}{2}} [q(x)]^{\frac{1+\alpha}{2}} d\mu(x)\right], \quad p,q \in \Omega; \tag{1.5}$$

$$D_B(p,q) := \int_\Gamma \sqrt{p(x)\,q(x)}\,d\mu(x), \quad p,q \in \Omega; \tag{1.6}$$

$$D_{Ha}(p,q) := \int_\Gamma \frac{2p(x)\,q(x)}{p(x)+q(x)} d\mu(x), \quad p,q \in \Omega; \tag{1.7}$$

$$D_J(p,q) := \int_\Gamma [p(x) - q(x)] \ln\left[\frac{p(x)}{q(x)}\right] d\mu(x), \quad p,q \in \Omega; \tag{1.8}$$

$$D_\Delta(p,q) := \int_\Gamma \frac{[p(x) - q(x)]^2}{p(x)+q(x)} d\mu(x), \quad p,q \in \Omega. \tag{1.9}$$

For other divergence measures, see the paper [5] by Kapur or the book on line [7] by Taneja. For a comprehensive collection of preprints available on line, see the RGMIA web site `http://rgmia.vu.edu.au/papersinfth.html`

Csiszár $f-$divergence is defined as follows [11]

$$D_f(p,q) := \int_\Gamma p(x)\, f\left[\frac{q(x)}{p(x)}\right] d\mu(x), \quad p,q \in \Omega, \tag{1.10}$$

where f is convex on $(0,\infty)$. It is assumed that $f(u)$ is zero and strictly convex at $u = 1$. By appropriately defining this convex function, various divergences are derived. All the above distances (1.1) – (1.9), are particular instances of Csiszár $f-$divergence. There are also many others which are not in this class (see for example [5] or [7]). For the basic properties of Csiszár $f-$divergence see [8]-[11].

In [12], Lin and Wong (see also [10]) introduced the following divergence

$$D_{LW}(p,q) := \int_\Gamma p(x) \log\left[\frac{p(x)}{\frac{1}{2}p(x) + \frac{1}{2}q(x)}\right] d\mu(x), \quad p,q \in \Omega. \tag{1.11}$$

This can be represented as follows, using the Kullback-Leibler divergence:

$$D_{LW}(p,q) = D_{KL}\left(p, \frac{1}{2}p + \frac{1}{2}q\right).$$

Lin and Wong have established the following inequalities

$$D_{LW}(p,q) \le \frac{1}{2} D_{KL}(p,q); \tag{1.12}$$

$$D_{LW}(p,q) + D_{LW}(q,p) \le D_v(p,q) \le 2; \tag{1.13}$$

$$D_{LW}(p,q) \le 1. \tag{1.14}$$

In [14], Shioya and Da-te improved (1.12) – (1.14) by showing that

$$D_{LW}(p,q) \le \frac{1}{2} D_v(p,q) \le 1.$$

For classical and new results in comparing different kinds of divergence measures, see the papers [1]-[14] where further references are given.

In [50], the authors introduced the following divergence measure

$$H_f(p,q;t) := \int_\chi p(x) f\left[\frac{tq(x)+(1-t)p(x)}{p(x)}\right] d\mu(x), \tag{1.15}$$

where $p,q\in Q$ and $t\in[0,1]$.

It is obvious that this measure can be represented in terms of Csiszár $f-$divergence, namely, we have the representation

$$H_f(p,q;t) = D_f(p, tq+(1-t)p) \tag{1.16}$$

for all $p,q\in Q$ and $t\in[0,1]$.

The following properties of $H_f(\cdot,\cdot;\cdot)$ hold (see [50]).

Theorem 1. *Assume that the mapping $f:[0,\infty)\to\mathbb{R}$ is convex and $p,q\in Q$. Then*

(i) *$H_f(p,q;\cdot)$ is convex on $[0,1]$;*
(ii) *We have the inequality*

$$H_f(p,q;t) \le D_f(p,q) \quad \textit{for all } t\in[0,1] \tag{1.17}$$

and the bounds

$$\inf_{t\in[0,1]} H_f(p,q;t) = H_f(p,q;0) = 0 \tag{1.18}$$

and

$$\sup_{t\in[0,1]} H_f(p,q;t) = H_f(p,q;1) = D_f(p,q); \tag{1.19}$$

(iii) *The mapping $H_f(p,q;\cdot)$ is monotonic nondecreasing on $[0,1]$.*

In the same paper [50], the authors introduced the following divergence

$$F_f(p,q;t) = \int_\chi\int_\chi p(x)p(y) f\left[t\cdot\frac{q(x)}{p(x)} + (1-t)\cdot\frac{q(y)}{p(y)}\right] d\mu(x)\,d\mu(y), \tag{1.20}$$

where $p,q\in\Omega$ and $t\in[0,1]$.

The properties of this mapping are embodied in the following theorem.

Theorem 2. *Under the assumptions of Theorem 1, we have*

(i) *$F_f(p,q;\cdot)$ is symmetrical about $\frac{1}{2}$, i.e.,*

$$F_f(p,q;t) = F_f(p,q;1-t) \quad \textit{for all } t\in[0,1]; \tag{1.21}$$

(ii) *$F_f(p,q;\cdot)$ is convex on $[0,1]$;*
(iii) *We have the bounds*

$$\sup_{t\in[0,1]} F_f(p,q;t) = F_f(p,q;0) = F_f(p,q;1) = D_f(p,q); \tag{1.22}$$

$$\begin{aligned}\inf_{t\in[0,1]} F_f(p,q;t) &= F_f\left(p,q;\frac{1}{2}\right) \\ &= \int_\chi\int_\chi p(x)p(y) f\left[\frac{q(x)p(y)+p(x)q(y)}{2p(x)p(y)}\right] d\mu(x)\,d\mu(y) \ge 0;\end{aligned} \tag{1.23}$$

(iv) *$F_f(p,q;\cdot)$ is nondecreasing on $\left[0,\frac{1}{2}\right]$ and nonincreasing on $\left[\frac{1}{2},1\right]$;*

(v) *We have the inequalities*

$$F_f(p,q;t) \geq \max\{H_f(p,q;t), H_f(p,q;1-t)\} \tag{1.24}$$

for all $t \in [0,1]$.

In this paper we point out some estimates for the divergence measures $F_f(\cdot,\cdot;\cdot)$ and $H_f(\cdot,\cdot;\cdot)$.

2. Some Estimates for $n-$Time Differentiable Mappings

We use the following lemma (see also [48]).

Lemma 1. *Let* $f : I \in \mathbb{R} \to \mathbb{R}$ *(I interval of* $\mathbb{R}$*) be such that* $f^{(n)}$ *is absolutely continuous on* I*. Then for all* $x, a \in \mathring{I}$ *(*$\mathring{I}$ *is the interior of* I*) we have the inequality*

$$\left| f(x) - f(a) - \sum_{k=1}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) \right| \tag{2.1}$$
$$\leq \begin{cases} \frac{1}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty |x-a|^{n+1} & \text{if } f^{(n+1)} \in L_\infty(I); \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha |x-a|^{n+\frac{1}{\beta}} & \text{if } f^{(n+1)} \in L_\alpha(I), \\ & \alpha > 1,\ \frac{1}{\alpha}+\frac{1}{\beta} = 1; \\ \frac{1}{n!} \left\| f^{(n+1)} \right\|_1 |x-a|^n, \end{cases}$$

where $\|\cdot\|_\alpha$ $(\alpha \in [1,\infty])$ *are the usual Lebesgue norms on* I*, i.e.,*

$$\|g\|_\alpha := \left(\int_I |g(x)|^\alpha dx \right)^{\frac{1}{\alpha}},\quad \alpha \geq 1$$
$$\|g\|_\infty := ess \sup_{x\in I} |g(x)|.$$

Proof. We start with the Taylor representation with the integral remainder

$$f(x) = f(a) + \sum_{k=1}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) + \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t)\,dt \tag{2.2}$$

for all $a, x \in \mathring{I}$.

Using the properties of modulus, we have

$$\left| f(x) - f(a) - \sum_{k=1}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) \right| \leq \frac{1}{n!} \left| \int_a^x |x-t|^n \left| f^{(n+1)}(t) \right| dt \right| =: M\left(f^{(n+1)}; a, x\right). \tag{2.3}$$

Obviously, we have

$$M\left(f^{(n+1)}; a, x\right) \leq ess \sup_{t\in I} \left| f^{(n+1)}(t) \right| \frac{1}{n!} \left| \int_a^x |x-t|^n dt \right| = \frac{1}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty |x-a|^{n+1} \tag{2.4}$$

for all $a, x \in \mathring{I}$.

In addition, by the use of the Hölder integral inequality, we have

$$
\begin{aligned}
(2.5)\qquad M\left(f^{(n+1)};a,x\right) &\le \frac{1}{n!}\left|\int_a^x |x-t|^{n\beta}\,dt\right|^{\frac{1}{\beta}}\left|\int_\alpha^\beta \left|f^{(n+1)}(t)\right|^\alpha dt\right|^{\frac{1}{\alpha}} \\
&= \frac{1}{n!}\left\|f^{(n+1)}\right\|_\alpha \left[\frac{(x-a)^{n\beta+1}}{n\beta+1}\right]^{\frac{1}{\beta}} \\
&= \frac{1}{n!\,(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha |x-a|^{n+\frac{1}{\beta}}, \\
\alpha &> 1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1,
\end{aligned}
$$

and finally

$$
\begin{aligned}
(2.6)\qquad M\left(f^{(n+1)};a,x\right) &\le \frac{1}{n!}|x-a|^n\left|\int_a^x \left|f^{(n+1)}(t)\right|dt\right| \\
&\le \frac{1}{n!}|x-a|^n\left\|f^{(n+1)}\right\|_1 .
\end{aligned}
$$

Now, by (2.3) - (2.6), we deduce the desired inequality (2.1). ∎

The following corollary will be useful in what follows.

Corollary 1. *Assume that f is as above and $a,b\in\mathring{I}$. Then for all $\lambda\in[0,1]$, we have the inequality:*

$$
(2.7)\qquad \left|f(\lambda b+(1-\lambda)a)-f(a)-\sum_{k=1}^n \frac{\lambda^k(b-a)^k}{k!}f^{(k)}(a)\right|
$$
$$
\le \begin{cases}
\frac{\lambda^{n+1}|b-a|^{n+1}}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty & \text{if } f^{(n+1)}\in L_\infty(I); \\
\frac{\lambda^{n+\frac{1}{\beta}}|b-a|^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\infty & \text{if } f^{(n+1)}\in L_\alpha(I), \\
 & \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\
\frac{\lambda^n|b-a|^n}{n!}\left\|f^{(n+1)}\right\|_1 .
\end{cases}
$$

We can now point out the following estimation result for the mapping $H_f(p,q;\cdot)$.

Theorem 3. *Assume that the mapping $f:[0,\infty)\to\mathbb{R}$ is such that $f^{(n)}$ is absolutely continuous on $[r,R]$, where $0\le r\le 1\le R<\infty$. If $p,q\in\Omega$ and*

$$
(2.8)\qquad r\le \frac{q(x)}{p(x)}\le R \quad a.e.\ on\ \chi,
$$

then we have the inequality

$$\left|H_f(p,q;t)-f(1)-\sum_{k=1}^{n}\frac{t^k f^{(k)}(1)}{k!}D_{\chi^k}(p,q)\right| \tag{2.9}$$

$$\leq\begin{cases}\frac{t^{n+1}\left\|f^{(n+1)}\right\|_{\infty}}{(n+1)!}D_{|\chi|^{n+1}}(p,q) & \text{if } f^{(n+1)}\in L_{\infty}[r,R]; \\ \frac{t^{n+\frac{1}{\beta}}\left\|f^{(n+1)}\right\|_{\alpha}}{n!(n\beta+1)^{\frac{1}{\beta}}}D_{|\chi|^{n+\frac{1}{\beta}}}(p,q) & \text{if } f^{(n+1)}\in L_{\alpha}[r,R], \\ & \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{t^{n}\left\|f^{(n+1)}\right\|_{1}}{n!}D_{|\chi|^{n}}(p,q), \end{cases}$$

$$\leq\begin{cases}\frac{t^{n+1}(R-r)^{n+1}}{(n+1)!}\left\|f^{(n+1)}\right\|_{\infty} \\ \frac{t^{n+\frac{1}{\beta}}(R-r)^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_{\alpha} \\ \frac{t^{n}(R-r)^{n}}{n!}\left\|f^{(n+1)}\right\|_{1}, \end{cases}$$

where

$$D_{\chi^k}(p,q):=\int_{\chi}\frac{(q(x)-p(x))^k}{p^{k-1}(x)}d\mu(x),\quad k=1,...$$

and

$$D_{|\chi|^r}(p,q):=\int_{\chi}\frac{|q(x)-p(x)|^r}{p^{r-1}(x)}d\mu(x),\quad r\geq 0$$

and the Lebesgue $\alpha-$*norms are taken on* $[r,R]$.

Proof. Apply inequality (2.1) for $\lambda=t\in[0,1]$, $b=\frac{q(x)}{p(x)}$, $x\in\chi$ and $a=1$, to get

$$\left|f\left(t\cdot\frac{q(x)}{p(x)}+(1-t)\right)-f(1)-\sum_{k=1}^{n}\frac{t^k\left(\frac{q(x)}{p(x)}-1\right)^k}{k!}f^{(k)}(1)\right| \tag{2.10}$$

$$\leq\begin{cases}\frac{t^{n+1}\left|\frac{q(x)}{p(x)}-1\right|^{n+1}}{(n+1)!}\left\|f^{(n+1)}\right\|_{\infty}; \\ \frac{t^{n+\frac{1}{\beta}}\left|\frac{q(x)}{p(x)}-1\right|^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_{\alpha}; \\ \frac{t^{n}\left|\frac{q(x)}{p(x)}-1\right|^{n}}{n!}\left\|f^{(n+1)}\right\|_{1}; \end{cases}\leq\begin{cases}\frac{t^{n+1}(R-r)^{n+1}}{(n+1)!}\left\|f^{(n+1)}\right\|_{\infty}; \\ \frac{t^{n+\frac{1}{\beta}}(R-r)^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_{\alpha}; \\ \frac{t^{n}(R-r)^{n}}{n!}\left\|f^{(n+1)}\right\|_{1}, \end{cases}$$

for a.e. $x\in\chi$.

If we multiply (2.10) by $p(x) \geq 0$, integrate on χ and use the properties of the integral, then we get

$$\Bigg|\int_{\chi} p(x) f\left(\frac{tq(x)+(1-t)p(x)}{p(x)}\right) d\mu(x) - f(1) - \sum_{k=1}^{n} \frac{t^k f^{(k)}(1)}{k!} \int_{\chi} \frac{(q(x)-p(x))^k}{p^{k-1}(x)} d\mu(x)\Bigg|$$

$$\leq \begin{cases} \frac{t^{n+1}\left\|f^{(n+1)}\right\|_{\infty}}{(n+1)!} \int_{\chi} \frac{|q(x)-p(x)|^{n+1}}{p^n(x)} d\mu(x); \\ \frac{t^{n+\frac{1}{\beta}}\left\|f^{(n+1)}\right\|_{\alpha}}{n!(n\beta+1)^{\frac{1}{\beta}}} \int_{\chi} \frac{|q(x)-p(x)|^{n+\frac{1}{\beta}}}{p^{n+\frac{1}{\beta}-1}(x)} d\mu(x); \\ \frac{t^{n}\left\|f^{(n+1)}\right\|_{1}}{n!} \int_{\chi} \frac{|q(x)-p(x)|^{n}}{p^{n-1}(x)} d\mu(x), \end{cases}$$

$$\leq \begin{cases} \frac{t^{n+1}(R-r)^{n+1}}{(n+1)!} \left\|f^{(n+1)}\right\|_{\infty} \\ \frac{t^{n+\frac{1}{\beta}}(R-r)^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\|f^{(n+1)}\right\|_{\alpha} \\ \frac{t^{n}(R-r)^{n}}{n!} \left\|f^{(n+1)}\right\|_{1}, \end{cases}$$

and the theorem is proved. ∎

Remark 1. *If $n=0$, then, basically, for an absolutely continuous mapping $f:[r,R] \subset [0,\infty) \to \mathbb{R}$, we have:*

$$|D_f(p, tq+(1-t)p) - f(1)| \tag{2.11}$$

$$\leq \begin{cases} t\|f'\|_{\infty} D_v(p,q) \\ t^{\frac{1}{\beta}}\|f'\|_{\alpha} D_{|\alpha|^{\frac{1}{\beta}}}(p,q) \\ \|f'\|_1 \end{cases} \leq \begin{cases} t\|f'\|_{\infty}(R-r) \\ t^{\frac{1}{\beta}}(R-r)^{\frac{1}{\beta}} \cdot \|f'\|_{\alpha} \\ \|f'\|_1 \end{cases}$$

for all $t \in [0,1]$, where $D_v(p,q) = \int_{\chi} |p(x)-q(x)| d\mu(x)$.
If $n=1$, and taking into account that $D_{\chi}(p,q)=0$, then by (2.5) we get for the mappings whose derivatives f' are absolutely continuous that

$$|D_f(p, tq+(1-t)p) - f(1)| \tag{2.12}$$

$$\leq \begin{cases} \frac{t^2\|f'\|_{\infty}}{2} D_{\chi^2}(p,q) & \text{if } f' \in L_{\infty}[r,R] \\ \frac{t^{\frac{\beta+1}{\beta}}\|f'\|_{\alpha}}{(\beta+1)^{\frac{1}{\beta}}} D_{|\chi|^{\frac{\beta+1}{\beta}}}(p,q) & \text{if } f' \in L_{\alpha}[r,R] \\ t\|f'\|_1 D_v(p,q) \end{cases}$$

for all $t \in [0,1]$.

Of course, if we assume that f is convex and normalised, then the left hand side of both (2.11) and (2.12) will become

$$0 \leq D_f\left(p, tq + (1-t)p\right)$$

and the inequalities (2.11) and (2.12) will provide some upper bounds for the mapping $H_f(p,q;t)$, $t \in [0,1]$.

Remark 2. *If we assume that f'' is absolutely continuous, then from (2.9) we obtain*

$$\left| H_f(p,q;t) - f(1) - \frac{t^2}{2} f''(1) D_{\chi^2}(p,q) \right| \tag{2.13}$$

$$\leq \begin{cases} \frac{t^3 \|f'''\|_\infty}{6} D_{|\chi|^3}(p,q) & \text{if } \; f''' \in L_\infty[r,R]; \\ \frac{t^{\frac{2\beta+1}{\beta}} \|f'''\|_\alpha}{2(2\beta+1)^{\frac{1}{\beta}}} D_{|\chi|^{2+\frac{1}{\beta}}}(p,q) & \text{if } \; f''' \in L_\alpha[r,R], \\ & \alpha > 1, \; \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \frac{t^2 \|f'''\|_1}{2} D_{|\chi|^2}(p,q), \end{cases}$$

which provides an approximation of $H_f(p,q;t)$ by a quadratic in t whose coefficient is dependent on the χ^2 -distance of p and q.

We also note that Theorem 3 contains, as a particular case (for $t = 1$), an approximation of the Csiszár $f-$divergence. Namely,

Corollary 2. *With the assumptions of Theorem 3, we have*

$$\left| D_f(p,q) - f(1) - \sum_{k=1}^{n} \frac{f^{(k)}(1)}{k!} D_{\chi^k}(p,q) \right| \tag{2.14}$$

$$\leq \begin{cases} \frac{\|f^{(n+1)}\|_\infty}{(n+1)!} D_{|\chi|^{n+1}}(p,q) \\ \frac{\|f^{(n+1)}\|_\alpha}{n!(n\beta+1)^{\frac{1}{\beta}}} D_{|\chi|^{n+\frac{1}{\beta}}}(p,q) \\ \frac{\|f^{(n+1)}\|_1}{n!} D_{|\chi|^n}(p,q) \end{cases} \leq \begin{cases} \frac{\|f^{(n+1)}\|_\infty}{(n+1)!} (R-r)^{n+1} \\ \frac{\|f^{(n+1)}\|_\alpha}{n!(n\beta+1)^{\frac{1}{\beta}}} (R-r)^{n+\frac{1}{\beta}} \\ \frac{\|f^{(n+1)}\|_1}{n!} (R-r)^n. \end{cases}$$

We also know that for $t = \frac{1}{2}$, we obtain the generalised Lin-Wong $f-$divergence

$$LW_f(p,q) := D_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right)$$

and so, from (2.9), we may state the following estimation for the Lin-Wong $f-$divergence.

Corollary 3. *With the assumptions of Theorem 3, we have*

$$\left|LW_f(p,q)-f(1)-\sum_{k=1}^{n}\frac{t^k f^{(k)}(1)}{2^k k!}D_{\chi^k}(p,q)\right| \tag{2.15}$$

$$\leq\begin{cases}\frac{\left\|f^{(n+1)}\right\|_\infty}{2^{n+1}(n+1)!}D_{|\chi|^{n+1}}(p,q) & \text{if } f^{(n+1)}\in L_\infty[r,R];\\ \frac{\left\|f^{(n+1)}\right\|_\alpha}{2^{n+\frac{1}{\beta}}n!(n\beta+1)^{\frac{1}{\beta}}}D_{|\chi|^{n+\frac{1}{\beta}}}(p,q) & \text{if } f^{(n+1)}\in L_\alpha[r,R],\\ & \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1;\\ \frac{\left\|f^{(n+1)}\right\|_1}{2^n n!}D_{|\chi|^n}(p,q).\end{cases}$$

Remark 3. *Similar particular cases for $n=0$, $n=1$ and $n=2$ may be stated, but we omit the details.*

The following theorem also holds.

Theorem 4. *Assume that the mapping $f:[0,\infty)\to\mathbb{R}$ is such that $f^{(n)}$ is absolutely continuous on $[r,R]$, where $0\leq r\leq 1\leq R<\infty$. If $p,q\in\Omega$ and*

$$r\leq\frac{q(x)}{p(x)}\leq R\quad \text{a.e. on } \chi, \tag{2.16}$$

then we have the inequality

$$\left|F_f(p,q;t)-D_f(p,q)-\sum_{k=1}^{n}\frac{t^k}{k!}D^{(*)}_{f^{(k)}}(p,q)\right| \tag{2.17}$$

$$\leq\begin{cases}\frac{t^{n+1}}{(n+1)!}D^{(*)}_{n+1}(p,q)\left\|f^{(n+1)}\right\|_\infty & \text{if } f^{(n+1)}\in L_\infty[r,R];\\ \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}}D^{(*)}_{n+\frac{1}{\beta}}(p,q)\left\|f^{(n+1)}\right\|_\alpha & \text{if } f^{(n+1)}\in L_\alpha[r,R],\\ & \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1;\\ \frac{t^n}{n!}D^{(*)}_n(p,q)\left\|f^{(n+1)}\right\|_1\end{cases}$$

$$\leq\begin{cases}\frac{t^{n+1}}{(n+1)!}(R-r)^{n+1}\left\|f^{(n+1)}\right\|_\infty & \text{if } f^{(n+1)}\in L_\infty[r,R];\\ \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}}(R-r)^{n+\frac{1}{\beta}}\left\|f^{(n+1)}\right\|_\alpha & \text{if } f^{(n+1)}\in L_\alpha[r,R],\\ & \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1;\\ \frac{t^n}{n!}(R-r)^n\left\|f^{(n+1)}\right\|_1,\end{cases}$$

where

$$D^{(*)}_{f^{(k)}}(p,q)=\int_\chi\int_\chi\frac{\left(\det\begin{bmatrix}p(y) & q(y)\\ p(x) & q(x)\end{bmatrix}\right)^k}{[p(x)]^{k-1}[p(y)]^{k-1}}f^{(k)}\left(\frac{q(y)}{p(y)}\right)d\mu(x)\,d\mu(y),\quad k=1,\ldots.$$

$$D_s^{(*)}(p,q)=\int_\chi\int_\chi \frac{\left|\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right|^s}{[p(x)]^{s-1}[p(y)]^{s-1}}d\mu(x)\,d\mu(y)\,,\quad s>0$$

and the $\alpha-$*norms are taken on* $[r,R]$.

Proof. We choose in Corollary 1, $b=\frac{q(x)}{p(x)}$, $a=\frac{q(y)}{p(y)}$, $x,y\in\chi$ to obtain

$$\left|f\left(t\cdot\frac{q(x)}{p(x)}+(1-t)\cdot\frac{q(y)}{p(y)}\right)-f\left(\frac{q(y)}{p(y)}\right)\right.$$
$$\left.-\sum_{k=1}^{n}\frac{t^k\left(\frac{q(x)}{p(x)}-\frac{q(y)}{p(y)}\right)^k}{k!}f^{(k)}\left(\frac{q(y)}{p(y)}\right)\right|$$
$$\leq\begin{cases}\frac{t^{n+1}}{(n+1)!}\left|\frac{q(x)}{p(x)}-\frac{q(y)}{p(y)}\right|^{n+1}\left\|f^{(n+1)}\right\|_\infty\\[2ex] \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}}\left|\frac{q(x)}{p(x)}-\frac{q(y)}{p(y)}\right|^{n+\frac{1}{\beta}}\left\|f^{(n+1)}\right\|_\alpha\\[2ex] \frac{t^n}{n!}\left|\frac{q(x)}{p(x)}-\frac{q(y)}{p(y)}\right|^{n}\left\|f^{(n+1)}\right\|_1\end{cases}$$

for all $x,y\in\chi$ and $t\in[0,1]$, which is clearly equivalent to

(2.18)
$$\left|f\left(\frac{tp(y)q(x)+(1-t)p(x)q(y)}{p(x)p(y)}\right)-f\left(\frac{q(y)}{p(y)}\right)\right.$$
$$\left.-\sum_{k=1}^{n}\frac{t^k}{k!}\cdot\frac{\left(\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right)^k}{[p(x)]^k[p(y)]^k}f^{(k)}\left(\frac{q(y)}{p(y)}\right)\right|$$
$$\leq\begin{cases}\frac{t^{n+1}}{(n+1)!}\cdot\frac{\left|\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right|^{n+1}}{[p(x)]^{n+1}[p(y)]^{n+1}}\left\|f^{(n+1)}\right\|_\infty\\[2ex] \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}}\cdot\frac{\left|\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right|^{n+\frac{1}{\beta}}}{[p(x)]^{n+\frac{1}{\beta}}[p(y)]^{n+\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha\\[2ex] \frac{t^n}{n!}\cdot\frac{\left|\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right|^{n}}{[p(x)]^{n}[p(y)]^{n}}\left\|f^{(n+1)}\right\|_1\end{cases}$$

for all $x,y\in\chi$ and $t\in[0,1]$.

If we multiply (2.18) by $p(x)p(y) \geq 0$ for $x, y \in \chi$, integrate over x and y on χ and use the properties of the integral, we obtain

$$\left| \int_\chi \int_\chi p(x)p(y) f\left(\frac{tp(y)q(x) + (1-t)p(x)q(y)}{p(x)p(y)}\right) d\mu(x)\, d\mu(y) \right.$$
$$- \int_\chi \int_\chi p(x)p(y) f\left(\frac{q(y)}{p(y)}\right) d\mu(x)\, d\mu(y)$$
$$\left. - \sum_{k=1}^{n} \frac{t^k}{k!} \int_\alpha \int_\chi \frac{\left(\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right)^k}{[p(x)]^{k-1}[p(y)]^{k-1}} f^{(k)}\left(\frac{q(y)}{p(y)}\right) d\mu(x)\, d\mu(y) \right|$$

$$\leq \begin{cases} \frac{t^{n+1}}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty \cdot \int_\chi \int_\chi \frac{\left|\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right|^{n+1}}{[p(x)]^n [p(y)]^n} d\mu(x)\, d\mu(y) \\ \\ \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha \cdot \int_\chi \int_\chi \frac{\left|\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right|^{n+\frac{1}{\beta}}}{[p(x)]^{n+\frac{1}{\beta}-1} [p(y)]^{n+\frac{1}{\beta}-1}} d\mu(x)\, d\mu(y) \\ \\ \frac{t^n}{n!} \left\| f^{(n+1)} \right\|_1 \cdot \int_\chi \int_\chi \frac{\left|\det\begin{bmatrix} p(y) & q(y) \\ p(x) & q(x)\end{bmatrix}\right|^{n}}{[p(x)]^{n-1} [p(y)]^{n-1}} d\mu(x)\, d\mu(y), \end{cases}$$

which is clearly equivalent to the first inequality in (2.17).

The second inequality is obvious by the fact that

$$\left| \frac{q(x)}{p(x)} - \frac{q(y)}{p(y)} \right| \leq R - r \quad \text{for all } x, y \in \chi.$$

The theorem is thus completely proved. ∎

References

[1] H. JEFFREYS, An invariant form for the prior probability in estimating problems, *Proc. Roy. Soc. London*, **186** A (1946), 453-461.

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

[3] A. RÉNYI, On measures of entropy and information, *Proc. Fourth Berkeley Symp. Math. Stat. and Prob., University of California Press*, **1** (1961), 547-561.

[4] J.H. HAVRDA and F. CHARVAT, Quantification method classification process: concept of structural α-entropy, *Kybernetika*, **3** (1967), 30-35.

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

[6] B.D. SHARMA and D.P. MITTAL, New non-additive measures of relative information, *Journ. Comb. Inf. Sys. Sci.,* **2** (4)(1977), 122-132.

[7] I. BURBEA and C.R. RAO, On the convexity of some divergence measures based on entropy function, *IEEE Trans. Inf. Th.,* **28** (3) (1982), 489-495.

[8] C.R. RAO, Diversity and dissimilarity coefficients: a unified approach, *Theoretic Population Biology*, **21** (1982), 24-43.

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

[10] I. CSISZÁR, Information-type measures of difference of probability distributions and indirect observations, *Studia Math. Hungarica,* **2** (1967), 299-318.
[11] I. CSISZÁR, On topological properties of $f-$divergences, *Studia Math. Hungarica,* **2** (1967), 329-339.
[12] 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.
[13] I. VAJDA, *Theory of Statistical Inference and Information,* Dordrecht-Boston, Kluwer Academic Publishers, 1989.
[14] M. MEI, The theory of genetic distance and evaluation of human races, *Japan J. Human Genetics*, **23** (1978), 341-369.
[15] A. SEN, *On Economic Inequality,* Oxford University Press, London 1973.
[16] H. THEIL, *Economics and Information Theory*, North-Holland, Amsterdam, 1967.
[17] H. THEIL, *Statistical Decomposition Analysis*, North-Holland, Amsterdam, 1972.
[18] E.C. PIELOU, *Ecological Diversity,* Wiley, New York, 1975.
[19] D.V. GOKHALE and S. KULLBACK, *Information in Contingency Tables*, New York, Merul Dekker, 1978.
[20] C.K. CHOW and C.N. LIN, Approximating discrete probability distributions with dependence trees, *IEEE Trans. Inf. Th.,* **14** (3) (1968), 462-467.
[21] D. KAZAKOS and T. COTSIDAS, A decision theory approach to the approximation of discrete probability densities, *IEEE Trans. Perform. Anal. Machine Intell.,* **1** (1980), 61-67.
[22] T.T. KADOTA and L.A. SHEPP, On the best finite set of linear observables for discriminating two Gaussian signals, *IEEE Trans. Inf. Th.,* **13** (1967), 288-294.
[23] T. KAILATH, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Comm. Technology.,* Vol COM-15 (1967), 52-60.
[24] M. BETH BASSAT, f-entropies, probability of error and feature selection, *Inform. Control*, **39** (1978), 227-242.
[25] C.H. CHEN, *Statistical Pattern Recognition,* Rocelle Park, New York, Hoyderc Book Co., 1973.
[26] V.A. VOLKONSKI and J. A. ROZANOV, Some limit theorems for random function -I, (English Trans.), *Theory Prob. Appl., (*USSR), **4** (1959), 178-197.
[27] M.S. PINSKER, Information and Information Stability of Random variables and processes, (in Russian), Moscow: Izv. Akad. Nouk, 1960.
[28] I. CSISZÁR, A note on Jensen's inequality, *Studia Sci. Math. Hung.,* **1** (1966), 185-188.
[29] H.P. McKEAN, JR., Speed of approach to equilibrium for Koc's caricature of a Maximilian gas, *Arch. Ration. Mech. Anal.,* **21** (1966), 343-367.
[30] J.H.B. KEMPERMAN, On the optimum note of transmitting information, *Ann. Math. Statist.,* **40** (1969), 2158-2177.
[31] S. KULLBACK, A lower bound for discrimination information in terms of variation, *IEEE Trans. Inf. Th.,* **13** (1967), 126-127.
[32] S. KULLBACK, Correction to a lower bound for discrimination information in terms of variation, *IEEE Trans. Inf. Th.,* **16** (1970), 771-773.
[33] I. VAJDA, Note on discrimination information and variation, *IEEE Trans. Inf. Th.,* **16** (1970), 771-773.
[34] G.T. TOUSSAINT, Sharper lower bounds for discrimination in terms of variation, *IEEE Trans. Inf. Th.,* **21** (1975), 99-100.
[35] F. TOPSOE, Some inequalities for information divergence and related measures of discrimination, *Res. Rep. Coll., RGMIA,* **2** (1) (1999), 85-98.
[36] L. LECAM, *Asymptotic Methods in Statistical Decision Theory,* New York: Springer, 1986.
[37] D. DACUNHA-CASTELLE, Ecole d'ete de Probability de Saint-Flour, III-1977, Berlin, Heidelberg: Springer 1978.
[38] C. KRAFT, Some conditions for consistency and uniform consistency of statistical procedures, *Univ. of California Pub. in Statistics*, **1** (1955), 125-142.
[39] S. KULLBACK and R.A. LEIBLER, On information and sufficiency, *Annals Math. Statist.,* **22** (1951), 79-86.
[40] E. HELLINGER, Neue Bergrüirdung du Theorie quadratisher Formerus von uneudlichvieleu Veränderlicher, *J. für reine and Augeur. Math.,* **36** (1909), 210-271.
[41] A. BHATTACHARYYA, On a measure of divergence between two statistical populations defined by their probability distributions, *Bull. Calcutta Math. Soc.,* **35** (1943), 99-109.

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

[43] I. CSISZÁR, A note on Jensen's inequality, *Studia Sci. Math. Hung.,* **1** (1966), 185-188.

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

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

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

[47] H. SHIOYA and T. DA-TE, A generalisation of Lin divergence and the derivative of a new information divergence, *Elec. and Comm. in Japan*, **78** (7) (1995), 37-40.

[48] S.S. DRAGOMIR, New estimation of the remainder in Taylor's formula using Grüss' type inequalities and applications, *Math. Ineq. and Appl.*, **2** (2) (1999), 183-193.

[49] S.S. DRAGOMIR, An improvement of the remainder estimate in the generalised Taylor's formula, *RGMIA Res. Rep. Coll.,* **3** (2000), *No. 1, Article 1.*

[50] N.S. BARNETT, P. CERONE and S.S. DRAGOMIR, Some new inequalities for Hermite-Hadamard difference in information theory, *submitted.*

[51] N.S. BARNETT, P. CERONE, S.S. DRAGOMIR and A. SOFO, Approximating Csiszár $f-$divergence by the use of Taylor's formula with integral remainder, *submitted.*

[52] M. MATIĆ, J. PEČARIĆ and N. UJEVIĆ, On new estimation of the remainder in generalised Taylor's formula, *Math. Ineq. and Appl.*, **2** (3) (1999), 343-361.

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

E-mail address: neil@matilda.vu.edu.au
URL: http://sci.vu.edu.au/staff/neilb.html

E-mail address: pc@matilda.vu.edu.au
URL: http://sci.vu.edu.au/staff/peterc.html

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

E-mail address: sofo@matilda.vu.edu.au