Some New Inequalities for Hermite-Hadamard Divergence in Information Theory

SOME NEW INEQUALITIES FOR HERMITE-HADAMARD DIVERGENCE IN INFORMATION THEORY

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

Abstract. In this paper we prove some new inequalities for Hermite-Hadamard divergence in Information Theory.

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 [7], Rao [8], Lin [9], Csiszár [10], Ali and Silvey [12], Vajda [13], Shioya and Da-te [40] and others (see for example [5] and the references therein).

These measures have been applied in a variety of fields such as: anthropology [8], genetics [14], finance, economics, and political science [15], [16], [17], biology [18], the analysis of contingency tables [19], approximation of probability distributions [20], [21], signal processing [22], [23] and pattern recognition [24], [25]. A number of these measures of distance are specific cases of 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.

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

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

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 [1], χ^2-*divergence* D_{χ^2}, $\alpha-$*divergence* D_α, *Bhattacharyya distance* D_B [2], *Harmonic distance* D_{Ha}, *Jeffreys distance* D_J [1], *triangular discrimination* D_Δ [35], etc... They are defined as follows:

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

Date: March 20, 2000.
1991 *Mathematics Subject Classification.* Primary94Xxx; Secondary 26D15.
Key words and phrases. Divergence measures in Information Theory.

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

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

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

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

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

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

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

For other divergence measures, see the paper [5] by Kapur or the book on line [6] 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 [10]

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

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 $f-$divergence. There are also many others that are not in this class (see for example [5] or [6]). For the basic properties of $f-$divergence see [7]-[10].

In [11], Lin and Wong (see also [9]) introduced the following divergence

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

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 [45], Shioya and Da-te improved $(1.12) - (1.14)$ by showing that

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

In the same paper [45], the authors introduced the generalised Lin-Wong $f-$divergence $D_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right)$ and the Hermite-Hadamard (HH) divergence

$$D_{HH}^{f}(p,q) := \int_{\chi} p(x) \frac{\int_1^{\frac{q(x)}{p(x)}} f(t)\,dt}{\frac{q(x)}{p(x)} - 1} d\mu(x), \quad p,q \in \Omega \tag{1.15}$$

and, by use of the Hermite-Hadamard inequality for convex functions, proved the following basic inequality

$$D_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right) \le D_{HH}^{f}(p,q) \le \frac{1}{2} D_f(p,q), \tag{1.16}$$

provided that f is convex and normalised, i.e., $f(1) = 0$.

In this paper we point out new inequalites for the $HH-$divergence, which also improve the above result (1.16).

For classical and new results in comparing different kinds of divergence measures, see the papers [1]-[45] where further references are given.

2. The Results

In the following, we assume everywhere that the mapping $f : (0, \infty) \to \mathbb{R}$ is convex and normalised.

The following result holds.

Theorem 1. *Let $p, q \in \not\leq$, then we have the inequality,*

$$\begin{aligned} & D_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right) \\ \le\ & \lambda D_f\left(p, p + \frac{\lambda}{2}(q-p)\right) + (1-\lambda) D_f\left(p, \frac{p+q}{2} + \frac{\lambda}{2}(q-p)\right) \\ \le\ & D_{HH}^{f}(p,q) \le \frac{1}{2}\left[D_f(p, (1-\lambda)p + \lambda q) + (1-\lambda) D_f(p,q)\right] \\ \le\ & \frac{1}{2} D_f(p,q), \end{aligned} \tag{2.1}$$

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

Proof. First, the following refinement of the Hermite-Hadamard inequality is proved.

$$\begin{aligned} & f\left(\frac{a+b}{2}\right) \\ \le\ & \lambda f\left(a + \lambda \cdot \frac{b-a}{2}\right) + (1-\lambda) f\left(\frac{a+b}{2} + \lambda \cdot \frac{b-a}{2}\right) \\ \le\ & \frac{1}{b-a} \int_a^b f(u)\,du \le \frac{1}{2}\left[f((1-\lambda)a + \lambda b) + \lambda f(a) + (1-\lambda) f(b)\right] \\ \le\ & \frac{f(a) + f(b)}{2} \end{aligned} \tag{2.2}$$

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

Applying the Hermite-Hadamard inequality on each subinterval $[a,(1-\lambda)a+\lambda b]$, $[(1-\lambda)a+\lambda b,b]$, we have,

$$\begin{aligned}
& f\left(\frac{a+(1-\lambda)a+\lambda b}{2}\right)\times[(1-\lambda)a+\lambda b-a] \\
\leq & \int_a^{(1-\lambda)a+\lambda b} f(u)\,du \\
\leq & \frac{f((1-\lambda)a+\lambda b)+f(a)}{2}\times[(1-\lambda)a+\lambda b-a]
\end{aligned}$$

and

$$\begin{aligned}
& f\left(\frac{(1-\lambda)a+\lambda b+b}{2}\right)\times[b-(1-\lambda)a-\lambda b] \\
\leq & \int_{(1-\lambda)a+\lambda b}^{b} f(u)\,du \\
\leq & \frac{f(b)+f((1-\lambda)a+\lambda b)}{2}\times[b-(1-\lambda)a-\lambda b],
\end{aligned}$$

which are clearly equivalent to

$$\begin{aligned}
\lambda f\left(a+\lambda\cdot\frac{b-a}{2}\right) &\leq \frac{1}{b-a}\int_a^{(1-\lambda)a+\lambda b} f(u)\,du \\
&\leq \frac{\lambda f((1-\lambda)a+\lambda b)+\lambda f(a)}{2}
\end{aligned}\tag{2.3}$$

and

$$\begin{aligned}
& (1-\lambda)f\left(\frac{a+b}{2}+\lambda\cdot\frac{b-a}{2}\right) \\
\leq & \frac{1}{b-a}\int_{(1-\lambda)a+\lambda b}^{b} f(u)\,du \\
\leq & \frac{(1-\lambda)f(b)+(1-\lambda)f((1-\lambda)a+\lambda b)}{2}
\end{aligned}\tag{2.4}$$

respectively.

Summing (2.3) and (2.4), we obtain the second and first inequality in (2.2).

By the convexity property, we obtain

$$\begin{aligned}
& \lambda f\left(a+\lambda\cdot\frac{b-a}{2}\right)+(1-\lambda)f\left(\frac{a+b}{2}+\lambda\cdot\frac{b-a}{2}\right) \\
\geq & f\left[\lambda\left(a+\lambda\cdot\frac{b-a}{2}\right)+(1-\lambda)\left(\frac{a+b}{2}+\lambda\cdot\frac{b-a}{2}\right)\right] \\
= & f\left(\frac{a+b}{2}\right)
\end{aligned}$$

and the first inequality in (2.1) is proved.

The latter inequality is obvious by the convexity property of f.

Now, if we choose $a = 1$ and $b = \frac{q(x)}{p(x)}$, $x \in \chi$, in (2.2) and multiply by $p(x) \geq 0$, $x \in \chi$, we get

$$
\begin{aligned}
& p(x) f\left(\frac{p(x)+q(x)}{2p(x)}\right) \\
\leq\; & \lambda p(x) f\left(\frac{p(x)+\lambda(q(x)-p(x))}{2p(x)}\right) \\
& +(1-\lambda) p(x) f\left(\frac{p(x)+q(x)}{2p(x)}+\frac{\lambda(q(x)-p(x))}{2p(x)}\right) \\
\leq\; & \frac{p^2(x)}{q(x)-p(x)} \int_1^{\frac{q(x)}{p(x)}} f(u)\, du \\
\leq\; & \frac{1}{2}\left[f\left(\frac{(1-\lambda) p(x)+\lambda q(x)}{p(x)}\right) p(x)+\lambda p(x) f(1)+(1-\lambda) p(x) f\left(\frac{q(x)}{p(x)}\right)\right] \\
\leq\; & \frac{p(x) f(1)+p(x) f\left(\frac{q(x)}{p(x)}\right)}{2}.
\end{aligned}
$$

Integrating on χ and taking into account the definition of $f-$divergence (1.10) and the Hermite-Hadamard divergence (1.15), we obtain (2.1). ∎

Remark 1. *If $\lambda = 0$ or $\lambda = 1$, then by (2.1), we obtain the inequality (1.16).*

Corollary 1. *Let $p, q \in \Omega$, then we have the inequality,*

$$
\begin{aligned}
D_f\left(p, \frac{p+q}{2}\right) \leq\; & \frac{1}{2}\left[D_f\left(p, \frac{3p+q}{4}\right)+D_f\left(p, \frac{p+3q}{4}\right)\right] \\
\leq\; & D_{HH}^f(p, q) \leq \frac{1}{2}\left[D_f\left(p, \frac{p+q}{2}\right)+\frac{1}{2} D_f(p, q)\right] \\
\leq\; & \frac{1}{2} D_f(p, q),
\end{aligned} \tag{2.5}
$$

which is obtained by taking $\lambda = \frac{1}{2}$ in (2.1).

Remark 2. *If we replace λ by $(1-\lambda)$ in (2.1), we have,*

$$
\begin{aligned}
& D_f\left(p, \frac{p+q}{2}\right) \\
\leq\; & (1-\lambda) D_f\left(p, \frac{p+q}{2}+\lambda(p-q)\right)+\lambda D_f\left(p, q+\lambda \frac{p-q}{2}\right) \\
\leq\; & D_{HH}^f(p, q) \leq \frac{1}{2}\left[D_f(p, \lambda p+(1-\lambda) q)+\lambda D_f(p, q)\right] \\
\leq\; & \frac{1}{2} D_f(p, q).
\end{aligned} \tag{2.6}
$$

Now, if we add (2.1) and (2.6) and divide by 2, we can state the following corollary.

Corollary 2. *Let $p, q \in \Omega$, then we have the inequality,*

$$
\begin{aligned}
(2.7) \qquad & D_f\left(p, \frac{p+q}{2}\right) \\
\leq\ & \lambda\left[D_f\left(p, p+\frac{\lambda}{2}(q-p)\right)+D_f\left(p, q+\frac{\lambda}{2}(p-q)\right)\right] \\
& +(1-\lambda)\left[D_f\left(p, \frac{p+q}{2}+\frac{\lambda}{2}(q-p)\right)+D_f\left(p, \frac{p+q}{2}+\frac{1}{2}(p-q)\right)\right] \\
\leq\ & D_{HH}^f(p, q) \\
\leq\ & \frac{1}{4}\left[D_f(p,(1-\lambda) p+\lambda q)+D_f(p, \lambda p+(1-\lambda) q)+D_f(p, q)\right] \\
\leq\ & \frac{1}{2} D_f(p, q),
\end{aligned}
$$

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

We also define the divergence.

$$
\begin{aligned}
(2.8) \qquad H_f(p, q; t) \quad : & = \int_\chi p(x) f\left[\frac{tq(x)+(1-t) p(x)}{p(x)}\right] d\mu(x) \\
& = D_f(p, tq+(1-t) p).
\end{aligned}
$$

Theorem 2. *Let $p, q \in \Omega$, then,*

(i) *$H_f(p, q; \cdot)$ is convex on $[0,1]$;*

(ii) *We have the bounds*

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

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

and the inequality

$$
(2.11) \qquad H_f(p, q; t) \leq t D_f(p, q) \quad \text{for all } t \in [0,1].
$$

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

Proof. (i) Let $t_1,\ t_2 \in [0,1]$ and $\alpha, \beta \in [0,1]$ with $\alpha+\beta=1$, then,

$$
\begin{aligned}
& H_f(p, q; \alpha t_1+\beta t_2) \\
=\ & \int_\chi p(x) f\left[\frac{(\alpha t_1+\beta t_2) q(x)+(1-\alpha t_1-\beta t_2) p(x)}{q(x)}\right] d\mu(x) \\
=\ & \int_\chi p(x) f\left[\alpha \cdot \frac{[t_1 q(x)+(1-t_1) p(x)]}{q(x)}+\beta \cdot \frac{[t_2 q(x)+(1-t_2) p(x)]}{q(x)}\right] d\mu(x) \\
\leq\ & \alpha \cdot \int_\chi p(x) f\left[\frac{t_1 q(x)+(1-t_1) p(x)}{q(x)}\right] d\mu(x) \\
& +\beta \cdot \int_\chi p(x) f\left[\frac{t_2 q(x)+(1-t_2) p(x)}{q(x)}\right] d\mu(x) \\
=\ & \alpha H_f(p, q, t_1)+\beta H_f(p, q, t_2)
\end{aligned}
$$

and convexity is proved.

(ii) Using Jensen's inequality, we have:

$$\begin{aligned} H_f(p,q,t) &\geq f\left[\int_{\mathcal{X}} p(x)\left[\frac{tq(x)+(1-t)p(x)}{q(x)}\right]d\mu(x)\right] \\ &= f\left[t\int_{\mathcal{X}} q(x)\,d\mu(x)+(1-t)\int_{\mathcal{X}} p(x)\,d\mu(x)\right] \\ &= f(1)=0=H_f(p,q,0). \end{aligned}$$

Also, by convexity of f, we have,

$$\begin{aligned} H_f(p,q,t) &\leq \int_{\mathcal{X}} p(x)\left[tf\left(\frac{q(x)}{p(x)}\right)+(1-t)f(1)\right]d\mu(x) \\ &\leq t\int_{\mathcal{X}} p(x)f\left(\frac{q(x)}{p(x)}\right)d\mu(x)+(1-t)f(1)\int_{\mathcal{X}} p(x)\,d\mu(x) \\ &= tD_f(p,q), \end{aligned}$$

and the statement (ii) is proved.

(iii) Let t_1, $t_2\in[0,1]$ with $t_2>t_1$. As $H_f(p,q;\cdot)$ is convex, then

$$\frac{H_f(p,q,t_2)-H_f(p,q,t_1)}{t_2-t_1}\geq\frac{H_f(p,q,t_1)-H_f(p,q,0)}{t_1-0}$$

and as

$$H_f(p,q,t_1)\geq H_f(p,q,0)=0,$$

we deduce that $H_f(p,q,t_1)\leq H_f(p,q,t_2)$, which proves the monotonicity of $H_f(p,q,\cdot)$.

∎

Remark 3. *If we write (2.11) in terms of* $1-t$ *rather than* t*, we obtain*

$$H_f(p,q,1-t)\leq(1-t)D_f(p,q),\quad t\in[0,1]. \tag{2.12}$$

Adding (2.11) and (2.12), we get,

$$H_f(p,q,t)+H_f(p,q,1-t)\leq D_f(p,q) \tag{2.13}$$

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

Remark 4. *For* $t\in\left[\frac{1}{2},1\right]$*, we have the inequality,*

$$D_f\left(p,\frac{1}{2}p+\frac{1}{2}q\right)\leq D_f(p,tq+(1-t)p)\leq tD_f(p,q), \tag{2.14}$$

which is similar to (1.13).

We can also define the divergence,

$$F_f(p,q;t):=\int_{\mathcal{X}}\int_{\mathcal{X}} p(x)p(y)f\left[t\cdot\frac{q(x)}{p(x)}+(1-t)\cdot\frac{q(y)}{p(y)}\right]d\mu(x)\,d\mu(y), \tag{2.15}$$

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

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

Theorem 3. *Let* $p,q\in\Omega$*, then,*

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

$$F_f(p,q;t)=F_f(p,q;1-t)\quad \text{for all } t\in[0,1]. \tag{2.16}$$

(ii) F *is convex on* $[0,1]$*;*

(iii) *We have the bounds:*

$$\sup_{t\in[0,1]} F_f(p,q;t) = F_f(p,q;0) = F_f(p,q;1) = D_f(p,q), \tag{2.17}$$

$$\begin{aligned} \inf_{t\in[0,1]} F_f(p,q;t) &= F_f\left(p,q;\frac{1}{2}\right) \\ &= \int_\chi\int_\chi p(x)\,p(y)\,f\left[\frac{q(x)\,p(y)+p(x)\,q(y)}{2p(x)\,q(y)}\right]d\mu(x)\,d\mu(y) \\ &\geq 0; \end{aligned} \tag{2.18}$$

(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 inequality:*

$$F_f(p,q;t) \geq \max\{H_f(p,q;t);H_f(p,q;1-t)\} \quad \textit{for all} \quad t\in[0,1]. \tag{2.19}$$

Proof. (i) Is obvious.

(ii) Follows by the convexity of f in a similar way to that in the proof of Theorem 2.

(iii) For all $x,y\in\chi$ we have:

$$f\left[t\cdot\frac{q(x)}{p(x)}+(1-t)\cdot\frac{q(y)}{p(y)}\right] \leq t\cdot f\left(\frac{q(x)}{p(x)}\right)+(1-t)\cdot f\left(\frac{q(y)}{p(y)}\right)$$

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

Multiplying by $p(x)\,p(y)\geq 0$ and integrating over χ^2, we write,

$$\begin{aligned} F_f(p,q;t) &\leq \int_\chi\int_\chi p(x)\,p(y)\left[t\cdot f\left(\frac{q(x)}{p(x)}\right)+(1-t)\cdot f\left(\frac{q(y)}{p(y)}\right)\right]d\mu(x)\,d\mu(y) \\ &= t\int_\chi p(y)\,d\mu(y)\int_\chi p(x)\,f\left(\frac{q(x)}{p(x)}\right)d\mu(x) \\ &\quad +(1-t)\int_\chi d\mu(x)\int_\chi p(y)\,f\left(\frac{q(y)}{p(y)}\right)d\mu(y) \\ &= t\cdot D_f(p,q)+(1-t)\cdot D_f(p,q) = D_f(p,q) \\ &= F_f(p,q;0) = F_f(p,q;1) \end{aligned}$$

and the bound (2.17) is proved.

Since f is convex, then for all $t\in[0,1]$ and $x,y\in\chi$, we have

$$\begin{aligned} &\frac{1}{2}\left\{f\left[t\cdot\frac{q(x)}{p(x)}+(1-t)\cdot\frac{q(y)}{p(y)}\right]+f\left[(1-t)\cdot\frac{q(x)}{p(x)}+t\cdot\frac{q(y)}{p(y)}\right]\right\} \\ \geq\; & f\left[\frac{1}{2}\left(\frac{q(x)}{p(x)}+\frac{q(y)}{p(y)}\right)\right]. \end{aligned}$$

Multiplying by $p(x)\,p(y)\geq 0$ and integrating over χ^2, we have,

$$\begin{aligned} &\frac{1}{2}\left[F_f(p,q;t)+F_f(p,q;1-t)\right] \\ \geq\; & \int_\chi\int_\chi p(x)\,p(y)\,f\left[\frac{1}{2}\left(\frac{q(x)}{p(x)}+\frac{q(y)}{p(y)}\right)\right]d\mu(x)\,d\mu(y) \end{aligned}$$

and the first part of (2.18) is proved.

Using Jensen's integral inequality, we may write:

$$\begin{aligned} & \int_{\chi}\int_{\chi} f\left[\frac{1}{2}\left(\frac{q(x)\,p(y)+p(x)\,q(y)}{p(x)\,q(y)}\right)\right] p(x)\,p(y)\,d\mu(x)\,d\mu(y) \\ \geq\ & f\left[\int_{\chi}\int_{\chi}\frac{1}{2}\left(\frac{q(x)\,p(y)+p(x)\,q(y)}{p(x)\,q(y)}\right) p(x)\,p(y)\,d\mu(x)\,d\mu(y)\right] \\ =\ & f\left[\frac{1}{2}\left[\int_{\chi} p(x)\,d\mu(x)\int_{\chi} p(y)\,d\mu(y)+\int_{\chi} q(x)\,d\mu(x)\int_{\chi} q(y)\,d\mu(y)\right]\right] \\ =\ & f(1)=0 \end{aligned}$$

and the second part of (2.18) is proved.

(iv) The mapping $F_f(p,q;\cdot)$ being convex on $[0,1]$, we may write, for $1 \geq t_2 > t_1 \geq \frac{1}{2}$, that,

$$\frac{F_f(p,q;t_2)-F_f(p,q;t_1)}{t_2-t_1} \geq \frac{F_f(p,q;t_1)-F_f\left(p,q;\frac{1}{2}\right)}{t_1-\frac{1}{2}}$$

and as

$$F_f(p,q;t_1) \geq F_f\left(p,q;\frac{1}{2}\right),\quad t_1 \geq \frac{1}{2},$$

we deduce that $F_f(p,q;t_2) \geq F_f(p,q;t_1)$, i.e., the mapping $F_f(p,q;\cdot)$ is monotonically nondecreasing on $\left[0,\frac{1}{2}\right]$.

Similarly, we can prove that $F_f(p,q;\cdot)$ is monotonically nonincreasing on $\left[0,\frac{1}{2}\right]$, and the statement (iv) is proved.

(v) Using Jensen's integral inequality, we have,

$$\begin{aligned} & \int_{\chi} p(y)\,f\left[t\cdot\frac{q(x)}{p(x)}+(1-t)\cdot\frac{q(y)}{p(y)}\right] d\mu(y) \\ \geq\ & f\left[\int_{\chi} p(y)\left[t\cdot\frac{q(x)}{p(x)}+(1-t)\cdot\frac{q(y)}{p(y)}\right] d\mu(y)\right] \\ =\ & f\left[t\cdot\frac{q(x)}{p(x)}\int_{\chi} p(y)\,d\mu(y)+(1-t)\cdot\int_{\chi} q(y)\,d\mu(y)\right] \\ =\ & f\left[t\cdot\frac{q(x)}{p(x)}+(1-t)\right]. \end{aligned}$$

Multiplying by $p(x) \geq 0$ and integrating over χ, we have,

$$\begin{aligned} F_f(p,q;t) &\geq \int_{\chi} p(x)\,f\left[t\cdot\frac{q(x)}{p(x)}+(1-t)\right] d\mu(x) \\ &= H_f(p,q;t), \end{aligned}$$

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

Now, as

$$F_f(p,q;1-t) \geq H_f(p,q;1-t)$$

and $F_f(p,q;t) = F_f(p,q;1-t)$ for all $t \in [0,1]$, the inequality (2.19) is 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, Marcel 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 and J. KÖRNER, *Information Theory: Coding Theorem for Discrete Memoryless Systems,* Academic Press, New York, 1981.
[44] J. LIN and S.K.M. WONG, A new directed divergence measure and its characterization, *Int. J. General Systems,* **17** (1990), 73-81.
[45] 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.

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://rgmia.vu.edu.au/cerone

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