A GENERALIZED $f-$DIVERGENCE FOR PROBABILITY VECTORS AND APPLICATIONS

S.S. DRAGOMIR

ABSTRACT. A generalized $f-$divergence for probability vectors and some fundamental inequalities are pointed out.

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 [31], Kullback and Leibler [42], Rényi [20], Havrda and Charvat [25], Kapur [44], Sharma and Mittal [5], Burbea and Rao [41], Rao [34], Lin [10], Csiszár [11], Ali and Silvey [52], Vajda [37], Shioya and Da-te [21] and others (see for example [44] and the references therein).

These measures have been applied in a variety of fields such as: anthropology [34], genetics [43], finance, economics, and political science [47], [48], [39], biology [19], the analysis of contingency tables [9], approximation of probability distributions [26], [23], signal processing [24], [3] and pattern recognition [8], [53]. 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 χ and the $\sigma-$finite measure μ are given. Consider the set of all probability densities on μ to be $\Omega := \left\{p | p : \chi \to \mathbb{R},\ p(x) \geq 0,\ \int p(x)\, d\mu(x) = 1\right\}$. The Kullback-Leibler divergence [42] is well known among the 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 [31], χ^2-*divergence* D_{χ^2}, $\alpha-$*divergence* D_α, *Bhattacharyya distance* D_B [42], *Harmonic distance* D_{Ha}, *Jeffreys distance* D_J [31], *triangular discrimination* D_Δ [33], 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.* Primary 94Xxx; Secondary 26D15.

Key words and phrases. A Generalized $f-$Divergence, Jensen's Inequality.

$$D_H(p,q) := \int_\chi \left[\sqrt{p(x)} - \sqrt{q(x)}\right]^2 d\mu(x), \quad p,q \in \Omega; \tag{1.3}$$

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

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

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

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

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

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

For other divergence measures, see the paper [25] by Kapur or the book on line [5] 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]

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

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 Csiszár $f-$divergence. There are also many others which are not in this class (see for example [44] or [5]). For the basic properties of Csiszár $f-$divergence see [41]-[11].

In [1], Lin and Wong (see also [10]) 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

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

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

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

In [37], Shioya and Da-te improved (1.12) − (1.14) by showing that

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

For classical and new results in comparing different kinds of divergence measures, see the papers [31]-[38] where further references are given.

2. The Generalized $f-$Divergence and Some Inequalities

Let $(p_1,...,p_n)$ and $(q_1,...,q_n)$ be such that $p_i, q_i \in \Omega$ $(i = 1,...,n)$. Define the $f-$divergence between two vectors by

$$\begin{aligned} &\bar{D}_f(p_1,...,p_n,q_1,...,q_n) \\ &:= \int_\chi ... \int_\chi p_1(x_1)...p_n(x_n) f\left[\frac{\frac{q_1(x_1)}{p_1(x_1)} + ... + \frac{q_n(x_n)}{p_n(x_n)}}{n}\right] d\mu(x_1)...d\mu(x_n) \end{aligned} \tag{2.1}$$

provided that the multiple integral exists and is finite and $f : [0,\infty) \to \mathbb{R}$ is a given mapping.

Usually, the mapping is assumed to be convex and normalised.

The following simple result holds.

Theorem 1. *Assume that the mapping $f : [0,\infty) \to \mathbb{R}$ is convex. Then we have the inequalities*

$$\begin{aligned} f(1) &\leq \bar{D}_f(p_1,...,p_{n+1},q_1,...,q_{n+1}) \\ &\leq \bar{D}_f(p_1,...,p_n,q_1,...,q_n) \leq ... \leq \bar{D}_f(p_1,p_2,q_1,q_2) \\ &\leq \bar{D}_f(p_1,q_1), \end{aligned} \tag{2.2}$$

provided that $p_i, q_i \in \Omega$, $i \in \mathbb{N}$ and $D_f(\cdot,\cdot)$ is the usual $f-$Csiszár divergence.

Proof. Using Jensen's inequality for multiple integrals, we obtain

$$\begin{aligned} &\int_\chi ... \int_\chi p_1(x_1)...p_n(x_n) f\left[\frac{1}{n}\left(\frac{q_1(x_1)}{p_1(x_1)} + ... + \frac{q_n(x_n)}{p_n(x_n)}\right)\right] d\mu(x_1)...d\mu(x_n) \\ \geq\ & f\left[\int_\chi ... \int_\chi p_1(x_1)...p_n(x_n) \cdot \frac{1}{n}\left(\frac{q_1(x_1)}{p_1(x_1)} + ... + \frac{q_n(x_n)}{p_n(x_n)}\right) d\mu(x_1)...d\mu(x_n)\right] \\ =\ & f(1) \end{aligned}$$

because a simple calculation shows that

$$\begin{aligned} &\frac{1}{n}\int_\chi ... \int_\chi p_1(x_1)...p_n(x_n)\left(\frac{q_1(x_1)}{p_1(x_1)} + ... + \frac{q_n(x_n)}{p_n(x_n)}\right) d\mu(x_1)...d\mu(x_n) \\ =\ & \frac{1}{n}\left[\int_\chi q_1(x_1)\,d\mu(x_1)\int_\chi p_2(x_2)\,d\mu(x_2)...\int_\chi p_n(x_n)\,d\mu(x_n)\right. \\ &\left. + \int_\chi p_1(x_1)\,d\mu(x_1)...\int_\chi q_n(x_n)\,d\mu(x_n)\right] = 1 \end{aligned}$$

and the first inequality in (2.2) is proved.

If we apply Jensen's inequality for $y_1, ..., y_{n+1}$, we may write

$$\frac{1}{n+1}\left[f\left(y_1\right)+...+f\left(y_{n+1}\right)\right] \geq f\left(\frac{y_1+...+y_{n+1}}{n+1}\right). \tag{2.3}$$

Choose

$$\begin{aligned} y_1 &= \frac{z_1+z_2+...+z_{n-1}+z_n}{n}, \\ y_2 &= \frac{z_2+z_3+...+z_n+z_{n+1}}{n}, \\ &\vdots \\ y_{n+1} &= \frac{z_{n+1}+z_1+...+z_{n-1}}{n}. \end{aligned}$$

Then

$$\begin{aligned} \frac{1}{n+1}\left(y_1+...+y_{n+1}\right) &= \frac{n\left(z_1+z_2+...+z_{n+1}\right)}{n\left(n+1\right)} \\ &= \frac{z_1+z_2+...+z_n+z_{n+1}}{n+1} \end{aligned}$$

and then, by (2.3) we have the inequality

$$\begin{aligned} &\frac{1}{n+1}\left[f\left(\frac{z_1+z_2+...+z_{n-1}+z_n}{n}\right)+...+f\left(\frac{z_{n+1}+z_1+...+z_{n-1}}{n}\right)\right] \\ \geq\; & f\left(\frac{z_1+z_2+...+z_n+z_{n+1}}{n+1}\right) \end{aligned} \tag{2.4}$$

for all $z_1, ..., z_{n+1} \in [0, \infty)$.

If we put in (2.4) $z_i = \frac{p_i(x_i)}{q_i(x_i)}$ $(i = 1, ..., n+1)$ and if we multiply with $p_1\left(x_1\right)...p_{n+1}\left(x_{n+1}\right) \geq 0$ and then integrate over χ^{n+1}, we may write

$$\begin{aligned} &\frac{1}{n+1}\left[\int_\chi ... \int_\chi p_1\left(x_1\right)...p_{n+1}\left(x_{n+1}\right)\right. \\ &\times f\left[\frac{1}{n}\left(\frac{p_1\left(x_1\right)}{q_1\left(x_1\right)}+...+\frac{p_n\left(x_n\right)}{q_n\left(x_n\right)}\right)\right] d\mu\left(x_1\right)...d\mu\left(x_{n+1}\right) \\ &... + \int_\chi ... \int_\chi p_1\left(x_1\right)...p_{n+1}\left(x_{n+1}\right) \\ &\left.\times\, f\left[\frac{1}{n}\left(\frac{p_{n+1}\left(x_{n+1}\right)}{q_{n+1}\left(x_{n+1}\right)}+\frac{p_1\left(x_1\right)}{q_1\left(x_1\right)}+...+\frac{p_{n-1}\left(x_{n-1}\right)}{q_{n-1}\left(x_{n-1}\right)}\right)\right] d\mu\left(x_1\right)...d\mu\left(x_{n+1}\right)\right] \\ \geq\; & \int_\chi ... \int_\chi p_1\left(x_1\right)...p_{n+1}\left(x_{n+1}\right) \\ &\times f\left[\frac{1}{n+1}\left(\frac{p_1\left(x_1\right)}{q_1\left(x_1\right)}+...+\frac{p_{n+1}\left(x_{n+1}\right)}{q_{n+1}\left(x_{n+1}\right)}\right)\right] d\mu\left(x_1\right)...d\mu\left(x_{n+1}\right) \\ =\; & \bar{D}_f\left(p_1, ..., p_{n+1}, q_1, ..., q_{n+1}\right) \end{aligned} \tag{2.5}$$

and as

$$\begin{aligned} & \frac{1}{n+1}\left[\int_{\chi}...\int_{\chi}p_{1}\left(x_{1}\right)...p_{n+1}\left(x_{n+1}\right)\right.\\ & \left.\times f\left[\frac{1}{n}\left(\frac{p_{1}\left(x_{1}\right)}{q_{1}\left(x_{1}\right)}+...+\frac{p_{n}\left(x_{n}\right)}{q_{n}\left(x_{n}\right)}\right)\right]d\mu\left(x_{1}\right)...d\mu\left(x_{n+1}\right)\right.\\ = & \ ...=\int_{\chi}...\int_{\chi}p_{1}\left(x_{1}\right)...p_{n+1}\left(x_{n+1}\right)\\ & \times f\left[\frac{1}{n}\left(\frac{p_{n+1}\left(x_{n+1}\right)}{q_{n+1}\left(x_{n+1}\right)}+...+\frac{p_{n-1}\left(x_{n-1}\right)}{q_{n-1}\left(x_{n-1}\right)}\right)\right]d\mu\left(x_{1}\right)...d\mu\left(x_{n+1}\right)\\ = & \ \bar{D}_{f}\left(p_{1},...,p_{n},q_{1},...,q_{n}\right), \end{aligned}$$

then, from (2.5), we deduce the second inequality for all n, and the theorem is thus proved. ∎

For $f:[0,\infty)\to\mathbb{R},\ f(x)=(x-1)^{2}$, we obtain the $\chi^{2}-$distance

$$D_{\chi^{2}}\left(p_{1},q_{1}\right)=\int_{\chi}p_{1}\left(x\right)\left(\frac{q_{1}\left(x\right)}{p_{1}\left(x\right)}-1\right)^{2}d\mu\left(x\right)=\int_{\alpha}\frac{q_{1}^{2}\left(x\right)-p_{1}^{2}\left(x\right)}{p_{1}\left(x\right)}d\mu\left(x\right),\quad p,q\in\Omega.$$

The following corollary holds.

Corollary 1. *Let $p_{i},q_{i}\in\Omega$ $(i\in\mathbb{N})$ and define the $\chi^{2}-$divergence by*

$$\begin{aligned} (2.6)\quad & \bar{D}_{\chi^{2}}\left(p_{1},...,p_{n};q_{1},...,q_{n}\right)\\ :\ = & \int_{\chi}...\int_{\chi}p_{1}\left(x_{1}\right)...p_{n}\left(x_{n}\right)\left[\frac{\left(\frac{q_{1}(x_{1})}{p_{1}(x_{1})}+...+\frac{q_{n}(x_{n})}{p_{n}(x_{n})}\right)}{n}-1\right]^{2}d\mu\left(x_{1}\right)...d\mu\left(x_{n}\right). \end{aligned}$$

Then

$$\bar{D}_{\chi^{2}}\left(p_{1},...,p_{n};q_{1},...,q_{n}\right)=\frac{1}{n^{2}}\sum_{i=1}^{n}D_{\chi^{2}}\left(p_{i},q_{i}\right) \tag{2.7}$$

and

$$\begin{aligned} (2.8)\qquad 0\ \leq & \ \bar{D}_{\chi^{2}}\left(p_{1},...,p_{n+1};q_{1},...,q_{n+1}\right)\leq\bar{D}_{\chi^{2}}\left(p_{1},...,p_{n};q_{1},...,q_{n}\right)\\ \leq & \ ...\leq\bar{D}_{\chi^{2}}\left(p_{1},p_{2},q_{1},q_{2}\right)\leq\bar{D}_{\chi^{2}}\left(p_{1},q_{1}\right) \end{aligned}$$

for all $n\in\mathbb{N}$, $n\geq1$.

Proof. We observe that

$$\begin{aligned}
& \bar{D}_{\chi^2}(p_1,...,p_n;q_1,...,q_n) \\
= & \frac{1}{n^2}\int_\chi ... \int_\chi p_1(x_1)...p_n(x_n) \\
& \times\left[\left(\frac{q_1(x_1)}{p_1(x_1)}-1\right)+...+\left(\frac{q_n(x_n)}{p_n(x_n)}-1\right)\right]^2 d\mu(x_1)...d\mu(x_n) \\
= & \frac{1}{n^2}\int_\chi ... \int_\chi p_1(x_1)...p_n(x_n)\left[\sum_{i=1}^n\left(\frac{q_i(x_i)}{p_i(x_i)}-1\right)^2\right. \\
& \left.+2\sum_{1\le i<j\le n}\left(\frac{q_i(x_i)}{p_i(x_i)}-1\right)\left(\frac{q_j(x_j)}{p_j(x_j)}-1\right)\right] d\mu(x_1)...d\mu(x_n) \\
= & \frac{1}{n^2}\sum_{i=1}^n\int_\chi ... \int_\chi p_1(x_1)...p_n(x_n)\left(\frac{q_i(x_i)}{p_i(x_i)}-1\right)^2 d\mu(x_1)...d\mu(x_n) \\
& +\frac{2}{n^2}\sum_{1\le i<j\le n}\int_\chi ... \int_\chi p_1(x_1)...p_n(x_n) \\
& \times\left(\frac{q_i(x_i)}{p_i(x_i)}-1\right)\left(\frac{q_j(x_j)}{p_j(x_j)}-1\right) d\mu(x_1)...d\mu(x_n).
\end{aligned}$$

However, for any i, we have

$$\begin{aligned}
& \int_\chi ... \int_\chi p_1(x_1)...p_n(x_n)\left(\frac{q_i(x_i)}{p_i(x_i)}-1\right)^2 d\mu(x_1)...d\mu(x_n) \\
= & \int_\chi p_1(x_1)\,d\mu(x_1)...\int_\chi p_i(x_i)\,d\mu(x_i)...\int_\chi p_n(x_n)\,d\mu(x_n) \\
= & \int_\chi p_i(x_i)\left(\frac{q_i(x_i)}{p_i(x_i)}-i\right)^2 d\mu(x_i)=\bar{D}_{\chi^2}(p_i,q_i)
\end{aligned}$$

and (for $i\neq j$)

$$\begin{aligned}
& \int_\chi ... \int_\chi p_1(x_1)...p_n(x_n)\left(\frac{q_i(x_i)}{p_i(x_i)}-1\right)\left(\frac{q_j(x_j)}{p_j(x_j)}-1\right) d\mu(x_1)...d\mu(x_n) \\
= & \int_\chi p_1(x_1)\,d\mu(x_1)...\int_\chi p_i(x_i)\left(\frac{q_i(x_i)}{p_i(x_i)}-1\right) d\mu(x_i) \\
& ...\int_\chi p_j(x_j)\left(\frac{q_j(x_j)}{p_j(x_j)}-1\right) d\mu(x_j) \\
& ...\int_\chi p_n(x_n)\,d\mu(x_n)=0
\end{aligned}$$

as

$$\int_\chi p_i(x_i)\left(\frac{q_i(x_i)}{p_i(x_i)}-1\right) d\mu(x_i)=\int_\chi p_j(x_j)\left(\frac{q_j(x_j)}{p_j(x_j)}-1\right) d\mu(x_j)=0$$

and then the representation (2.7) is proved.

The sequence of inequalities in (2.8) follows by (2.2) and we omit the details. ∎

For $f:[0,\infty)\to\mathbb{R}$, $f(x)=e^{x}$, define the exp-*divergence* by

$$D_{\exp(\cdot)}(p_1,q_1):=\int_{\chi}p_1(x)\exp\left(\frac{q_1(x)}{p_1(x)}\right)d\mu(x),\quad p,q\in\Omega. \tag{2.9}$$

The following corollary holds.

Corollary 2. *Let $p_i,q_i\in\Omega$ $(i\in\mathbb{N})$ and define the* exp-*divergence by*

$$\begin{aligned}&\bar{D}_{\exp}(p_1,...,p_n;q_1,...,q_n)\\ :&=\int_{\chi}...\int_{\chi}p_1(x_1)...p_n(x_n)\exp\left(\frac{\frac{q_1(x_1)}{p_1(x_1)}+...+\frac{q_n(x_n)}{p_n(x_n)}}{n}\right)d\mu(x_1)...d\mu(x_n).\end{aligned} \tag{2.10}$$

Then

$$\bar{D}_{\exp}(p_1,...,p_n;q_1,...,q_n)=\prod_{i=1}^{n}D_{[\exp(\cdot)]^{\frac{1}{n}}}(p_i,q_i) \tag{2.11}$$

and

$$\begin{aligned}1&\le\bar{D}_{\exp}(p_1,...,p_{n+1};q_1,...,q_{n+1})\le\bar{D}_{\exp}(p_1,...,p_n;q_1,...,q_n)\\ &\le...\le\bar{D}_{\exp}(p_1,p_2,q_1,q_2)\le D_{\exp}(p_1,q_1),\end{aligned} \tag{2.12}$$

for all $n\geq1$.

Proof. We observe that

$$\begin{aligned}&\bar{D}_{\exp}(p_1,...,p_n;q_1,...,q_n)\\ =&\int_{\chi}...\int_{\chi}p_1(x_1)...p_n(x_n)\exp\left[\frac{q_1(x_1)}{p_1(x_1)}\right]...\exp\left[\frac{q_n(x_n)}{p_n(x_n)}\right]d\mu(x_1)...d\mu(x_n)\\ =&\prod_{i=1}^{n}\int_{\chi}p_i(x_i)\exp\left[\frac{q_i(x_i)}{np_i(x_i)}\right]d\mu(x_i)\\ =&\prod_{i=1}^{n}\int_{\chi}p_i(x_i)\left(\exp\frac{q_i(x_i)}{p_i(x_i)}\right)^{\frac{1}{n}}d\mu(x_i)=\prod_{i=1}^{n}D_{[\exp(\cdot)]^{\frac{1}{n}}}(p_i,q_i)\end{aligned}$$

and the identity (2.11) is proved.

The sequence of inequalities (2.12) follows by (2.2). ∎

References

[1] 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.

[2] N.S. BARNETT, S.S. DRAGOMIR and A. SOFO, Better bounds for an inequality of the Ostrowski type with applications, *RGMIA Research Report Collection,* **3** (1) (2000), Article 1.

[3] M. BETH BASSAT, f-entropies, probability of error and feature selection, *Inform. Control*, **39** (1978), 227-242.

[4] A. BHATTACHARYYA, On a measure of divergence between two statistical populations defined by their probability distributions, *Bull. Calcutta Math. Soc.,* **35** (1943), 99-109.

[5] 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.

[6] P. CERONE and S.S. DRAGOMIR, Midpoint type rules from an inequalities point of view, *RGMIA Research Report Collect.,* **2** (1999), No. 7, Article 1 (`http://rgmia.vu.edu.au/v2n7.html`)

[7] P. CERONE and S.S. DRAGOMIR, Trapezoidal type rules from an inequalities point of view, *RGMIA Research Report Collect.,* **2** (1999), No. 6, Article 8 (`http://rgmia.vu.edu.au/v2n6.html`)

[8] C.H. CHEN, *Statistical Pattern Recognition,* Rocelle Park, New York, Hoyderc Book Co., 1973.

[9] C.K. CHOW and C.N. LIN, Approximating discrete probability distributions with dependence trees, *IEEE Trans. Inf. Th.,* **14** (3) (1968), 462-467.

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

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

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

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

[16] D. DACUNHA-CASTELLE, *Ecole d'ete de Probability de Saint-Flour*, III-1977, Berlin, Heidelberg: Springer 1978.

[17] S.S. DRAGOMIR and S. MABIZELA, Some error estimates in the trapezoidal quadrature rule, *RGMIA Research Report Collect.,* **2** (1999), No. 5, Article 6 (`http://rgmia.vu.edu.au/v2n5.html`)

[18] S.S. DRAGOMIR and S. WANG, A new inequality of Ostrowski-Grüss type and its applications to the estimation of error bounds for special means and for some numerical quadrature rules, *Comp. Math. Appl.,* **33** (1997), No. 11, 15-20.

[19] D.V. GOKHALE and S. KULLBACK, *Information in Contingency Tables*, New York, Marcel Dekker, 1978.

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

[21] E. HELLINGER, Neue Bergrüirdung du Theorie quadratisher Formerus von uneudlichvieleu Veränderlicher, *J. für reine and Augeur. Math.,* **36** (1909), 210-271.

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

[23] 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.

[24] T. KAILATH, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Comm. Technology.,* Vol COM-15 (1967), 52-60.

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

[26] 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.

[27] J.H.B. KEMPERMAN, On the optimum note of transmitting information, *Ann. Math. Statist.,* **40** (1969), 2158-2177.

[28] C. KRAFT, Some conditions for consistency and uniform consistency of statistical procedures, *Univ. of California Pub. in Statistics*, **1** (1955), 125-142.

[29] S. KULLBACK, A lower bound for discrimination information in terms of variation, *IEEE Trans. Inf. Th.,* **13** (1967), 126-127.

[30] S. KULLBACK, Correction to a lower bound for discrimination information in terms of variation, *IEEE Trans. Inf. Th.,* **16** (1970), 771-773.

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

[32] S. KULLBACK and R.A. LEIBLER, On information and sufficiency, *Annals Math. Statist.,* **22** (1951), 79-86.

[33] L. LECAM, *Asymptotic Methods in Statistical Decision Theory,* New York: Springer, 1986.

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

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

[36] 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.
[37] M. MEI, The theory of genetic distance and evaluation of human races, *Japan J. Human Genetics*, **23** (1978), 341-369.
[38] C.E.M. PEARCE, J. PEČARIĆ, N. UJEVIĆ and S. VAROŠANEC, Generaliasations of some inequalities of Ostrowski-Grüss type, *Math. Ineq. & Appl.,* **3** (2000), No. 1, 25-34.
[39] E.C. PIELOU, *Ecological Diversity,* Wiley, New York, 1975.
[40] M.S. PINSKER, Information and Information Stability of Random variables and processes, (in Russian), Moscow: Izv. Akad. Nouk, 1960.
[41] C.R. RAO, Diversity and dissimilarity coefficients: a unified approach, *Theoretic Population Biology*, **21** (1982), 24-43.
[42] A. RÉNYI, On measures of entropy and information, *Proc. Fourth Berkeley Symp. Math. Stat. and Prob., University of California Press*, **1** (1961), 547-561.
[43] A. SEN, *On Economic Inequality,* Oxford University Press, London 1973.
[44] B.D. SHARMA and D.P. MITTAL, New non-additive measures of relative information, *Journ. Comb. Inf. Sys. Sci.,* **2** (4)(1977), 122-132.
[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.
[46] I. J. TANEJA, *Generalised Information Measures and their Applications* (http://www.mtm.ufsc.br/~taneja/bhtml/bhtml.html).
[47] H. THEIL, *Economics and Information Theory*, North-Holland, Amsterdam, 1967.
[48] H. THEIL, *Statistical Decomposition Analysis*, North-Holland, Amsterdam, 1972.
[49] F. TOPSOE, Some inequalities for information divergence and related measures of discrimination, *Res. Rep. Coll., RGMIA,* **2** (1) (1999), 85-98.
[50] G.T. TOUSSAINT, Sharper lower bounds for discrimination in terms of variation, *IEEE Trans. Inf. Th.,* **21** (1975), 99-100.
[51] I. VAJDA, Note on discrimination information and variation, *IEEE Trans. Inf. Th.,* **16** (1970), 771-773.
[52] I. VAJDA, *Theory of Statistical Inference and Information,* Dordrecht-Boston, Kluwer Academic Publishers, 1989.
[53] V.A. VOLKONSKI and J. A. ROZANOV, Some limit theorems for random function -I, (English Trans.), *Theory Prob. Appl., (*USSR), **4** (1959), 178-197.

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

E-mail address: sever.dragomir@vu.edu.au

URL: http://rgmia.vu.edu.au/SSDragomirWeb.html