Stolarsky and Gini Divergence Measures in Information Theory

STOLARSKY AND GINI DIVERGENCE MEASURES IN INFORMATION THEORY

P. CERONE AND S.S. DRAGOMIR

Abstract. In this paper we introduce the concepts of Stolarsky and Gini divergence measures and establish a number of basic properties. Some comparison results in the same class or between different classes are also given.

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 [14], Kullback and Leibler [18], Rényi [27], Havrda and Charvat [12], Kapur [15], Sharma and Mittal [28], Burbea and Rao [4], Rao [26], Lin [20], Csiszár [6], Ali and Silvey [1], Vajda [35], Shioya and Da-te [29] and others (see for example [15] and the references therein).

Assume that a set χ and the $\sigma-$finite measure μ are given. Consider the set of all probability densities on χ to be $\Omega := \left\{p|p : \chi \to \mathbb{R},\, p(t) \geq 0,\ \int_\chi p(t)\, d\mu(t) = 1\right\}$. The Kullback-Leibler divergence [18] is well known among the information divergences. It is defined as:

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

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 [13], χ^2-*divergence* D_{χ^2}, $\alpha-$*divergence* D_α, *Bhattacharyya distance* D_B [2], *Harmonic distance* D_{Ha}, *Jeffrey's distance* D_J [14], *triangular discrimination* D_Δ [33], etc... They are defined as follows:

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

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

Date: April 01, 2004.
1991 *Mathematics Subject Classification.* Primary 94Xxx; Secondary 26D15.
Key words and phrases. f-Divergence, Divergence measures in Information Theory, Stolarsky means, Gini means.

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

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

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

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

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

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

For other divergence measures, see the paper [15] by Kapur or the book on line [32] by Taneja. For a comprehensive collection of preprints available on line, see the RGMIA web site `http://rgmia.vu.edu.au/papersinfth.html.`

The $f-$divergence is defined as follows [6] :

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

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 which are not in this class (see for example [15] or [32]). For the basic properties of the $f-$divergence see [7]-[9].

For classical and new results in comparing different kinds of divergence measures, see the papers [1]-[36] where further references are given.

2. L_p- Divergence Measures

We define the $p-$*Logarithmic means* by (see [3, p. 346])

$$L_p(a,b) = \begin{cases} \left[\frac{b^{p+1}-a^{p+1}}{(p+1)(b-a)}\right]^{\frac{1}{p}}, & p\neq -1,0, \\ \frac{b-a}{\ln b - \ln a}, & p=-1, \qquad a\neq b,\ a,b>0 \\ \frac{1}{e}\left(\frac{b^b}{a^a}\right)^{\frac{1}{b-a}}, & p=0, \end{cases} \tag{2.1}$$

$$L_p(a,a) = a.$$

Where convenient, $L_{-1}(a,b)$ $-$*the logarithmic mean*, will be written as just $L(a,b)$. The case $p=0$ is also called the *identric mean*, i.e., $L_0(a,b)$ and will be denoted by $I(a,b)$. Of course, we will also define $L_\infty(a,b)=\max\{a,b\}$ and $L_{-\infty}=\min\{a,b\}$ to complete the special cases.

It is easily checked that the above definitions are consistent in the sense that $\lim_{p\to 0} L_p(a,b) = I(a,b)$ and $\lim_{p\to\pm\infty} L_p(a,b) = L_{\pm\infty}(a,b)$.

Following [10], we define the $p-$logarithmic divergence measure, or simply the L_p-*divergence measure,* by

$$\begin{aligned}(2.2)\qquad D_{L_p}(q,r) &: = \int_{\chi} L_p(q(t),r(t))\,d\mu(t)\\ &= \begin{cases} \int_{\chi}\left[\frac{[q(t)]^{p+1}-[r(t)]^{p+1}}{(p+1)(q(t)-r(t))}\right]^{\frac{1}{p}} d\mu(t), & \text{if } p\neq -1,0,\\ \int_{\chi}\left[\frac{q(t)-r(t)}{\ln q(t)-\ln r(t)}\right] d\mu(t), & \text{if } p=-1,\\ \frac{1}{e}\int_{\chi}\left[\frac{[q(t)]^{q(t)}}{[r(t)]^{r(t)}}\right]^{\frac{1}{q(t)-r(t)}} d\mu(t), & \text{if } p=0,\end{cases}\end{aligned}$$

$$(2.3)\qquad D_{+\infty}(q,r) : = \int_{\chi}\max\{q(t),r(t)\}\,d\mu(t),$$

$$(2.4)\qquad D_{-\infty}(q,r) : = \int_{\chi}\min\{q(t),r(t)\}\,d\mu(t),$$

for any $q,r\in\Omega$. We observe that

$$(2.5)\qquad D_{+\infty}(q,r) = \int_{\chi}\frac{q(t)+r(t)+|q(t)-r(t)|}{2}d\mu(t) = 1+\frac{1}{2}D_v(q,r)$$

and similarly,

$$(2.6)\qquad D_{-\infty}(q,r) = 1-\frac{1}{2}D_v(q,r).$$

Since $L_p(a,b) = L_p(b,a)$ for all $a,b>0$ and $p\in[-\infty,\infty]$, we can conclude that the L_p-*divergence measures are symmetrical.*

Now, if we consider the continuous mappings (which are not necessarily convex)

$$\begin{aligned}(2.7)\qquad f_p(t) &: = L_p(t,1)\\ &= \begin{cases}\left[\frac{t^{p+1}-1}{(p+1)(t-1)}\right]^{\frac{1}{p}}, & t\in(0,1)\cup(1,\infty),\ p\neq -1,0;\\ \frac{t-1}{\ln t}, & t\in(0,1)\cup(1,\infty),\ p=-1;\\ \frac{1}{e}t^{\frac{t}{t-1}}, & t\in(0,1)\cup(1,\infty),\ p=0;\\ 1 & \text{if } t=1,\end{cases}\\ f_{+\infty}(t) &= 1+\frac{1}{2}|t-1|,\\ f_{-\infty}(t) &= 1-\frac{1}{2}|t-1|,\end{aligned}$$

and since $L_p(a,b) = aL_p\left(1, \frac{b}{a}\right)$ for all $a, b > 0$ and $p \in [-\infty, \infty]$, we deduce that

$$D_{f_p}(q,r) = \int_\chi q(t) f_p\left(\frac{r(t)}{q(t)}\right) d\mu(t) = \int_\chi q(t) L_p\left(\frac{r(t)}{q(t)}, 1\right) d\mu(t) = \int_\chi L_p(r(t), q(t)) d\mu(t) = D_{L_p}(q,r), \tag{2.8}$$

for all $q, r \in \Omega$, which shows that the L_p-*divergence measure can be interpreted as an* $f-$*divergence,* for $f = f_p$, that are not necessarily convex.

The following fundamental theorem regarding the position of the L_p-divergence measures has been obtained in [10].

Theorem 1. *For any $q, r \in \Omega$, we have the inequality*

$$1 - \frac{1}{2}D_v(r,q) \le D_{L_u}(r,q) \le D_{L_s}(r,q) \le 1 + \frac{1}{2}D_v(r,q) \tag{2.9}$$

for all $-\infty \le u < s \le \infty$.
In particular, we have

$$1 - \frac{1}{2}D_v(p,q) \le D_{[HG^2]^{\frac{1}{3}}}(p,q) \le D_B(p,q) \le D_L(p,q) \le \frac{1}{2} + \frac{1}{2}D_B(p,q) \le D_I(p,q) \le 1, \tag{2.10}$$

where

$$D_L(r,q) := \int_\chi \left[\frac{r(t) - q(t)}{\ln r(t) - \ln q(t)}\right] d\mu(t) \text{ is the Logarithmic divergence,}$$

$$D_I(r,q) = \frac{1}{e}\int_\chi \left[\frac{[r(t)]^{r(t)}}{[q(t)]^{q(t)}}\right]^{\frac{1}{r(t)-q(t)}} d\mu(t) \text{ is the Identric divergence,}$$

$$D_B(p,q) = \int_\chi \sqrt{r(x)\,q(x)}\,d\mu(x) \text{ (Bhattacharyya distance)}$$

and

$$D_{[HG^2]^{\frac{1}{3}}}(p,q) := \sqrt[3]{2}\int_\chi \frac{r^{\frac{2}{3}}(t)\,q^{\frac{2}{3}}(t)}{[r(t) + q(t)]^{\frac{1}{3}}} d\mu(t).$$

Remark 1. *From (2.9), we can conclude the following inequality for the L_p-divergence measure in terms of the variational distance*

$$|D_{L_s}(r,q) - 1| \le \frac{1}{2}D_v(r,q), \; r, q \in \Omega \tag{2.11}$$

for all $s \in [-\infty, \infty]$. The constant $\frac{1}{2}$ is sharp.

3. $\alpha-$Power Divergence Measures

For $r\in\mathbb{R}$, we define the $\alpha-$*th power mean* of the positive numbers a,b by (see [3, p. 133])

$$M^{[\alpha]}(a,b):=\begin{cases}\left(\frac{a^{\alpha}+b^{\alpha}}{2}\right)^{\frac{1}{\alpha}} & \text{if} \quad \alpha\neq 0,\ \alpha\neq\pm\infty;\\ \sqrt{ab} & \text{if} \quad \alpha=0;\\ \max\{a,b\} & \text{if} \quad \alpha=+\infty;\\ \min\{a,b\} & \text{if} \quad \alpha=-\infty.\end{cases} \tag{3.1}$$

Following [10], we define the $\alpha-$*power divergence measure* by

$$\begin{aligned} D_{M^{[\alpha]}}(p,q) \quad &: =\int_{\chi}M^{[\alpha]}(p(t),q(t))\,d\mu(t) \\ &= \begin{cases}\int_{\chi}\left[\frac{p^{\alpha}(t)+q^{\alpha}(t)}{2}\right]^{\frac{1}{\alpha}}d\mu(t) & \text{if} \quad \alpha\neq 0,\ \alpha\neq\pm\infty;\\ \int_{\chi}\sqrt{p(t)\,q(t)}\,d\mu(t) & \text{if} \quad \alpha=0;\\ 1+\frac{1}{2}D_{v}(p,q) & \text{if} \quad \alpha=+\infty;\\ 1-\frac{1}{2}D_{v}(p,q) & \text{if} \quad \alpha=-\infty.\end{cases}\end{aligned} \tag{3.2}$$

Since $M^{[\alpha]}(a,b)=M^{[\alpha]}(b,a)$ for all $a,b>0$ and $\alpha\in[-\infty,\infty]$, we can conclude that the $\alpha-$*power divergences are symmetrical*. Now, if we consider the continuous functions (that are not necessarily convex)

$$f_{\alpha}(t):=M^{[\alpha]}(t,1)=\begin{cases}\left[\frac{t^{\alpha}+1}{2}\right]^{\frac{1}{\alpha}} & \text{if} \quad \alpha\neq 0,\ \alpha\neq\pm\infty;\\ \sqrt{t} & \text{if} \quad \alpha=0,\ t\in(0,\infty);\\ 1+\frac{1}{2}|t-1| & \text{if} \quad \alpha=+\infty;\\ 1-\frac{1}{2}|t-1| & \text{if} \quad \alpha=-\infty\end{cases} \tag{3.3}$$

and taking into account that $M^{[\alpha]}(a,b)=aM^{[\alpha]}\left(1,\frac{b}{a}\right)$, we deduce that

$$\begin{aligned} D_{f_{\alpha}}(p,q) &= \int_{\chi}p(t)\,f_{\alpha}\left(\frac{q(t)}{p(t)}\right)d\mu(t) \\ &= \int_{\chi}p(t)\,M^{[\alpha]}\left(\frac{q(t)}{p(t)},1\right)d\mu(t) \\ &= \int_{\chi}M^{[\alpha]}(p(t),q(t))\,d\mu(t)=D_{M^{[\alpha]}}(p,q),\end{aligned} \tag{3.4}$$

for all $p,r\in\Omega$, which shows that the $\alpha-$*power divergence measures can be interpreted as $f-$divergences*, for $f=f_{\alpha}$.

The following theorem concerning the location of the $\alpha-$power divergence measure has been obtained in [10].

Theorem 2. *For any $p,q\in\Omega$, we have:*

$$1-\frac{1}{2}D_v(p,q)\le D_{M^{[\alpha]}}(p,q)\le D_{M^{[\beta]}}(p,q)\le 1+\frac{1}{2}D_v(p,q) \tag{3.5}$$

for $-\infty\le\alpha<\beta\le\infty$.
In particular, we have

$$1-\frac{1}{2}D_v(p,q)\le D_{Ha}(p,q)\le D_B(p,q)\le 1+\frac{1}{2}D_v(p,q), \tag{3.6}$$

where $D_{Ha}(p,q)$ is the Harmonic divergence and $D_B(p,q)$ is the Bhattacharyya distance.

Remark 2. *From (3.5), we may conclude the following inequalities for the α-power divergence measures in terms of the variational distance*

$$\left|D_{M^{[\alpha]}}(p,q)-1\right|\le\frac{1}{2}D_v(p,q) \tag{3.7}$$

for any $p,q\in\Omega$ and $\alpha\in[-\infty,\infty]$ and the constant $\frac{1}{2}$ is sharp.

The following relationship between the power divergence and the generalized logarithmic divergence holds (see [10]).

Theorem 3. *For any $p,q\in\Omega$, we have:*

$$D_{M^{[r_1]}}(p,q)\le D_{L_r}(p,q)\le D_{M^{[r_2]}}(p,q), \tag{3.8}$$

where r_1 are defined by

$$r_1:=\begin{cases}\min\left\{\frac{r+2}{3},\ r\cdot\frac{\ln 2}{\ln r+1}\right\}, & \text{if } r>-1,\ r\neq 0;\\ \min\left\{\frac{2}{3},\ \ln 2\right\}, & \text{if } r=0;\\ \min\left\{\frac{r+2}{3},\ 0\right\}, & \text{if } r\le -1;\end{cases}$$

and r_2, as defined above, but with max *instead of* min.

Using the above means, we can imagine more divergences that can be constructed by the use of different contributions of these means. For example, we can define for $p,q\in\Omega$

$$\begin{aligned} D_{(AG)^{\frac{1}{2}}}(p,q) &:= \int_\chi\sqrt{A(p(t),q(t))\,G(p(t),q(t))}\,d\mu(t);\\ D_{(LI)^{\frac{1}{2}}}(p,q) &:= \int_\chi\sqrt{L(p(t),q(t))\,I(p(t),q(t))}\,d\mu(t)\end{aligned}$$

or even

$$D_{(GI)^{\frac{1}{2}}}(p,q):=\int_\chi\sqrt{G(p(t),q(t))\,I(p(t),q(t))}\,d\mu(t).$$

Using Alzer's result for means (see for instance [3, p. 350]), we may state the following theorem concerning the above divergence measures (see [10]).

Theorem 4. *For any $p,q\in\Omega$, we have*

$$D_{(AG)^{\frac{1}{2}}}(p,q) < D_{(LI)^{\frac{1}{2}}}(p,q)<D_{M^{[\frac{1}{2}]}}(p,q), \tag{3.9}$$

$$D_L(p,q)+D_I(p,q) < 1+B(p,q), \tag{3.10}$$

$$D_{(GI)^{\frac{1}{2}}}(p,q) < D_I(p,q)<\frac{1}{2}\left[B(p,q)+D_I(p,q)\right]. \tag{3.11}$$

4. Stolarsky Means & Divergence Measures

Let $a,b\in\mathbb{R}$ and let $x,y>0$. The *Stolarsky mean* $D_{(a,b)}(x,y)$ of order (a,b) of x and y $(x\neq y)$ is defined as

$$D_{(a,b)}(x,y)=\begin{cases}\left[\frac{b(x^a-y^a)}{a(x^b-y^b)}\right]^{\frac{1}{b-a}} & \text{if}\quad (a-b)\,ab\neq 0;\\ \exp\left(-\frac{1}{a}+\frac{x^a\ln x-y^a\ln y}{x^a-y^a}\right) & \text{if}\quad a=b\neq 0;\\ \left[\frac{x^a-y^a}{a(\ln x-\ln y)}\right]^{\frac{1}{a}} & \text{if}\quad a\neq 0,\quad b=0;\\ \sqrt{xy} & \text{if}\quad a=b=0,\end{cases}\tag{4.1}$$

with $D_{(a,b)}(x,x)=0$ (see [30]).

The following properties of the Stolarsky mean follow immediately from its definition

(i) $D_{(a,b)}(\cdot,\cdot)$ is symmetric in its parameters, i.e., $D_{(a,b)}(\cdot,\cdot)=D_{(b,a)}(\cdot,\cdot)$;
(ii) $D_{(\cdot,\cdot)}(x,y)$ is symmetric in the variables x and y, i.e., $D_{(\cdot,\cdot)}(x,y)=D_{(\cdot,\cdot)}(y,x)$;
(iii) $D_{(a,b)}(x,y)$ is a homogeneous function of order one in its variables, i.e.,

$$D_{(a,b)}(tx,ty)=tD_{(a,b)}(x,y),\quad t>0.$$

It can be proved that $D_{(\cdot,\cdot)}(\cdot,\cdot)$ is an infinitely many times differentiable function on $\mathbb{R}^2\times\mathbb{R}_+^2$, where $\mathbb{R}_+$ denotes the set of positive reals. This calss of means contains several particular means. For instance,

$$D_{(a,a)}(x,y)=I_a(x,y)\tag{4.2}$$

is the identric mean of order a, while

$$D_{(a,0)}(x,y)=L_a(x,y)\tag{4.3}$$

is the logarithmic mean of order a. Also,

$$D_{(2a,a)}(x,y)=M_a(x,y)\tag{4.4}$$

is the power mean of order a. For the inequality connecting the logarithmic means of order one and the power means, see [5], [22] and the references therein.

Let

$$\mu(x,y):=\begin{cases}\frac{|x|-|y|}{x-y} & \text{if}\quad x\neq y;\\ \operatorname{sgn}(x) & \text{if}\quad x=y.\end{cases}$$

The definition of the logarithmic mean is extended to the domain $x,y\geq 0$ by

$$L(x,y):=\begin{cases}\frac{x-y}{\ln x-\ln y} & \text{if}\quad x,y>0,\ x\neq y;\\ 0 & \text{if}\quad x,y=0.\end{cases}$$

The following comparison theorem for the Stolarsky mean is due to Leach-Sholander [19] and Páles [24].

Theorem 5. *Let $a,b,c,d\in\mathbb{R}$. Then the comparison inequality*

$$D_{(a,b)}(x,y)\leq D_{(c,d)}(x,y)\tag{4.5}$$

holds true for all $x,y>0$ *if and only if* $a+b\leq c+d$ *and*

$$\begin{aligned} L(a,b) &\leq L(c,d) \quad \textit{if} \ \ 0<\min(a,b,c,d),\\ \mu(a,b) &\leq \mu(c,d) \quad \textit{if} \ \ \min(a,b,c,d)<0<\max(a,b,c,d),\\ -L(-a,-b) &\leq -L(-c,-d) \quad \textit{if} \ \ \max(a,b,c,d)\leq 0. \end{aligned}$$

Using Theorem 5, we can prove that $D_{(\cdot,b)}(x,y)$ are increasing functions for all $b\in\mathbb{R}$ and all $x,y>0$.

We define the $(a,b)-$*Stolarsky Divergence Measures* by

$$\begin{aligned} &D_{(a,b)}(p,q) \\ &:= \int_{\chi} D_{(a,b)}(p(t),q(t))\,d\mu(t) \\ &= \begin{cases} \int_{\chi}\left[\frac{b(p^a(t)-q^a(t))}{a(p^b(t)-q^b(t))}\right]d\mu(t) & \text{if} \quad (a-b)ab\neq 0;\\ \int_{\chi}\exp\left(-\frac{1}{a}+\frac{p^a(t)\ln p(t)-q^a(t)\ln q(t)}{p^a(t)-q^a(t)}\right)d\mu(t) & \text{if} \quad a=b\neq 0;\\ \int_{\chi}\left[\frac{p^a(t)-q^a(t)}{a(\ln p(t)-\ln q(t))}\right]^{\frac{1}{a}}d\mu(t) & \text{if} \quad a\neq 0, \quad b=0;\\ \int_{\chi}\sqrt{p(t)\,q(t)}\,d\mu(t) & \text{if} \quad a=b=0, \end{cases} \end{aligned} \tag{4.6}$$

where $p,q\in\Omega$.

Since Stolarsky's means are symmetrical for $(a,b)\in\mathbb{R}^2$, we may conclude that $D_{(a,b)}(p,q)=D_{(a,b)}(q,p)$ for each $p,q\in\Omega$, i.e., the Stolarsky divergence measures are also symmetrical.

Now, if we consider the functions (that are not necessarily convex)

$$f_{(a,b)}(t) := D_{(a,b)}(t,1) = \begin{cases} \left[\frac{b(t^a-1)}{a(t^b-1)}\right]^{\frac{1}{b-a}} & \text{if} \quad (a-b)ab\neq 0;\\ \exp\left(-\frac{1}{a}+\frac{t^a\ln t}{t^a-1}\right) & \text{if} \quad a=b\neq 0;\\ \left[\frac{t^a-1}{a\ln t}\right]^{\frac{1}{a}} & \text{if} \quad a\neq 0, \quad b=0;\\ \sqrt{t} & \text{if} \quad a=b=0, \end{cases}$$

and taking into account that for any $x,y>0$

$$D_{(a,b)}(x,y)=yD_{(a,b)}\left(\frac{x}{y},1\right)$$

we deduce the equality

$$\begin{aligned} D_{f_{(a,b)}}(p,q) &= \int_{\chi} p(t)\, f_{(a,b)}\left(\frac{q(t)}{p(t)}\right) d\mu(t) \\ &= \int_{\chi} p(t)\, D_{(a,b)}\left(\frac{q(t)}{p(t)}, 1\right) d\mu(t) \\ &= \int_{\chi} D_{(a,b)}(q(t), p(t))\, d\mu(t) \\ &= D_{(a,b)}(p,q) \end{aligned}$$

for each $p, q \in \Omega$, which shows that the $(a,b)-$Stolarsky divergence measures can be interpreted as $f-$divergences for $f = f_{(a,b)}$. Note that, in general, $f_{(a,b)}$ are not convex functions.

Using the comparison theorem, we may state the following result for Stolarsky divergence measures.

Theorem 6. *Let $a, b, c, d \in \mathbb{R}$. If $a + b \leq c + d$ and*

$$\begin{aligned} L(a,b) &\leq L(c,d) \quad \text{if} \quad 0 < \min(a,b,c,d), \\ \mu(a,b) &\leq \mu(c,d) \quad \text{if} \quad \min(a,b,c,d) < 0 < \max(a,b,c,d), \\ -L(-a,-b) &\leq -L(-c,-d) \quad \text{if} \quad \max(a,b,c,d) \leq 0, \end{aligned}$$

then we have the inequalities:

$$1 - \frac{1}{2} D_v(p,q) \leq D_{(a,b)}(p,q) \leq D_{(c,d)}(p,q) \leq 1 + \frac{1}{2} D_v(p,q). \tag{4.7}$$

The first and last inequalities in (4.7) follow by the fact that for any pair of real numbers (a,b), one has

$$\min(x,y) \leq D_{(a,b)}(x,y) \leq \max(x,y)$$

for any $x, y > 0$.

5. Gini Means & Divergence Measures

In 1938, C. Gini [11] introduced the following means

$$S_{(a,b)}(x,y) = \begin{cases} \left(\frac{x^a + y^a}{x^b + y^b}\right)^{\frac{1}{a-b}} & \text{if} \quad a \neq b; \\ \exp\left(\frac{x^a \ln x + y^a \ln y}{x^a + y^a}\right) & \text{if} \quad a = b \neq 0; \\ \sqrt{xy} & \text{if} \quad a = b = 0, \end{cases} \tag{5.1}$$

where $a, b \in \mathbb{R}$.

Clearly, these means are symmetric in their parameters and variables, also, they are homogeneous of order one in their variables. It is worth mentioning that

$$S_{(a,0)}(x,y) = M_{[a]}(x,y).$$

In 1988, Zs. Páles [25] proved the following comparison theorem.

Theorem 7. *Let $a,b,c,d \in \mathbb{R}$. Then the comparison inequality*

$$S_{(a,b)}(x,y) \leq S_{(c,d)}(x,y) \tag{5.2}$$

is valid for all $x,y>0$ if and only if $a+b \leq c+d$ and

$$\begin{aligned} \min(a,b) &\leq \min(c,d) \quad \text{if } \ 0<\min(a,b,c,d), \\ \mu(a,b) &\leq \mu(c,d) \quad \text{if } \ \min(a,b,c,d)<0<\max(a,b,c,d), \\ \max(a,b) &\leq \max(c,d) \quad \text{if } \ \max(a,b,c,d) \leq 0. \end{aligned}$$

Using Theorem 7 one can prove that $S_{(\cdot,b)}(x,y)$ are increasing functions for all $b \in \mathbb{R}$ and all $x,y>0$.

We define the $(a,b)-$*Gini Divergence Measures* by

$$\begin{aligned} S_{(a,b)}(p,q) &:= \int_{\chi} S_{(a,b)}(p(t),q(t))\,d\mu(t) \\ &= \begin{cases} \int_{\chi} \left(\frac{p^a(t)+q^a(t)}{p^b(t)+q^b(t)}\right)^{\frac{1}{a-b}} d\mu(t) & \text{if} \quad a \neq b; \\ \int_{\chi} \exp\left(\frac{p^a(t)\ln p(t)+q^a(t)\ln q(t)}{p^a(t)+q^a(t)}\right) d\mu(t) & \text{if} \quad a=b\neq 0; \\ \int_{\chi} \sqrt{p(t)\,q(t)}\,d\mu(t) & \text{if} \quad a=b=0, \end{cases} \end{aligned} \tag{5.3}$$

where $a,b \in \mathbb{R}$.

It is obvious that the $(a,b)-$Gini divergence measures are symmetrical. If we consider the functions

$$g_{(a,b)}(t) := S_{(a,b)}(t,1) = \begin{cases} \left(\frac{t^a+1}{t^b+1}\right)^{\frac{1}{a-b}} & \text{if} \quad a\neq b; \\ \exp\left(\frac{t^a \ln t}{t^a+1}\right) & \text{if} \quad a=b\neq 0; \\ \sqrt{t} & \text{if} \quad a=b=0, \end{cases}$$

and taking into account that for any $x,y>0$

$$S_{(a,b)}(x,y) = yS_{(a,b)}\left(\frac{x}{y},1\right)$$

we deduce the equality

$$D_{g_{(a,b)}}(p,q) = D_{(a,b)}(p,q)$$

for any $p,q \in \Omega$, which shows that the $(a,b)-$Gini divergence measures can be interpreted as $f-$divergences for $f=g_{(a,b)}$. Note that in general $g_{(a,b)}$ are not convex functions.

Using the comparison theorem for Gini means, we may state the following result concerning the Gini divergence measures.

Theorem 8. *Let $a,b,c,d \in \mathbb{R}$. If $a+b \leq c+d$ and*

$$\begin{aligned} L(a,b) &\leq L(c,d) \quad \text{if } \ 0<\min(a,b,c,d), \\ \mu(a,b) &\leq \mu(c,d) \quad \text{if } \ \min(a,b,c,d)<0<\max(a,b,c,d), \\ -L(-a,-b) &\leq -L(-c,-d) \quad \text{if } \ \max(a,b,c,d) \leq 0. \end{aligned}$$

then we have the inequalities

$$1-\frac{1}{2}D_{v}\left(p,q\right)\leq S_{(a,b)}\left(p,q\right)\leq S_{(c,d)}\left(p,q\right)\leq 1+\frac{1}{2}D_{v}\left(p,q\right). \tag{5.4}$$

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

[3] P.S. BULLEN, D.S MITRINOVIĆ and P.M. VASIĆ, *Means and Their Inequalities*, Kluwer Academic Publishers, 1988.

[4] 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.

[5] B.C. CARLSON, The logarithmic mean, *Amer. Math. Monthly,* **79** (1972), 615-618.

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

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

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

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

[10] S.S. DRAGOMIR, On the p-logarithmic and α- power divergence measures in information theory, *PanAmerican Math. Journal*, **13**(3) (2003), 1-10. (Preprint on line: *RGMIA Res. Rep. Coll.*, **5**(2002), Article 25, `http://rgmia.vu.edu.au/v5(E).html`).

[11] C. GINI, Di una formula delle medic, Metron, **13** (1938), 3-22.

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

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

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

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

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

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

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

[19] E.B. LEACH and M.C. SHOLANDER, Multi-variable extended mean values, *J. Math. Anal. Appl.,* **104** (1984), 390-407.

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

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

[22] T.P. LIN, The power mean and the logarithmic mean, *Amer. Math. Monthly,* **81** (1974), 271-281.

[23] E. NEUMAN, The weighted logarithmic mean, *J. Math. Anal. Appl.,* **188** (1994), 885-900.

[24] Zs. PÁLES, Inequalities for differences of powers, *J. Math. Anal. Appl.,* **131** (1988), 271-281.

[25] Zs. PÁLES, Inequalities for sums of powers, *J. Math. Anal. Appl.,* **131** (1988), 265-270.

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

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

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

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

[30] K.B. STOLARSKY, Generalizations of the logarithmic mean, *Math. Mag.,* **41** (1975), 87-92.

[31] K.B. STOLARSKY, The power and generalized logarithmic means, *Amer. Math. Monthly,* **87** (1980), 545-548.

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

[33] F. TOPSOE, Some inequalities for information divergence and related measures of discrimination, *Res. Rep. Coll., RGMIA,* **2** (1) (1999), 85-98.

[34] G.T. TOUSSAINT, Sharper lower bounds for discrimination in terms of variation, *IEEE Trans. Inf. Th.,* **21** (1975), 99-100.

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

[36] I. VAJDA, Note on discrimination information and variation, *IEEE Trans. Inf. Th.,* **16** (1970), 771-773.

School of Computer Science and Mathematics, Victoria University of Technology, PO Box 14428, MCMC 8001, VICTORIA, Australia.

E-mail address: pc@csm.vu.edu.au

URL: http://rgmia.vu.edu.au/cerone/index.html

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

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