Some Inequalities for f-Divergence Measures Generated by 2n-Convex Functions

SOME INEQUALITIES FOR $f-$DIVERGENCE MEASURES GENERATED BY $2n-$CONVEX FUNCTIONS

S.S. DRAGOMIR AND S. KOUMANDOS

Abstract. A double Jensen type inequality for $2n-$convex functions is obtained and applied to establish upper and lower bounds for the $f-$divergence measure in Information Theory. Some particular inequalities of interest are stated as well.

1. Introduction

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space satisfying $|\mathcal{A}| > 2$ and μ a $\sigma-$finite measure on Ω. Let $\mathcal{P}$ be the set of all probability measures on the measurable space $(\Omega, \mathcal{A})$ which are absolutely continuous with respect to μ. For $P, Q \in \mathcal{P}$, let $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ denote the *Radon-Nikodym* derivatives of P and Q with respect to μ. Two probability measures $P, Q \in \mathcal{P}$ are said to be *orthogonal* and we denote this by $Q \perp P$ if

$$P(\{q = 0\}) = Q(\{p = 0\}) = 1.$$

Let $f : [0, \infty) \to (-\infty, \infty]$ be a convex function that is continuous at 0, i.e., $f(0) = \lim_{u \downarrow 0} f(u)$.

In 1963, I. Csiszár [3] introduced the concept of $f-$divergence as follows.

Definition 1. *Let $P, Q \in \mathcal{P}$. Then*

$$I_f(Q, P) := \int_\Omega p(s) f\left[\frac{q(s)}{p(s)}\right] d\mu(s), \tag{1.1}$$

is called the $f-$divergence of the probability distributions Q and P.

We observe that the integrand in (1.1) is undefined when $p(s) = 0$. We can overcome this problem by postulating for f as above that

$$0 f\left[\frac{q(s)}{0}\right] = q(s) \lim_{u \downarrow 0} \left[u f\left(\frac{1}{u}\right)\right], \quad s \in \Omega. \tag{1.2}$$

We recall now some important classes of $f-$divergences that play a key role in various problems in Information Theory and Statistics.

A. *The class of $\chi-$divergences.* The $f-$divergences in this class are generated by the family of functions $f_\alpha(u) := |u - 1|^\alpha$, $u \in [0, \infty)$, $\alpha \in [1, \infty)$. They have the form:

$$I_{f_\alpha}(Q, P) = \int_\Omega p^{1-\alpha}(s) |q(s) - p(s)|^\alpha d\mu(s). \tag{1.3}$$

Date: December 10, 2007.

2000 *Mathematics Subject Classification.* 94Axx, 26D15, 26D10.

Key words and phrases. $f-$divergence measure, $2n-$convexity, Convex functions, Absolutely monotonic and completely monotonic functions, Analytic inequalities.

From this class only the parameter $\alpha = 1$ provides a distance in the metric sense, namely, the *total variation distance*

$$V(Q,P) := \int_{\Omega} |q(s) - p(s)|\, d\mu(s).$$

The most prominent special case in this class is however the Karl Pearson χ^2-divergence

$$I_{\chi^2}(Q,P) = \int_{\Omega} \frac{(q(s) - p(s))^2}{p(s)} d\mu(s) = \int_{\Omega} \frac{q^2(s)}{p(s)} d\mu(s) - 1. \tag{1.4}$$

B. *The Dichotomy class.* This class is generated by the family of functions $g_\alpha : [0,\infty) \to \mathbb{R}$ where

$$g_\alpha(u) := \begin{cases} u - 1 - \log u & \text{for } \alpha = 0, \\ \frac{1}{\alpha(1-\alpha)}[\alpha u + 1 - \alpha - u^\alpha] & \text{for } \alpha \in \mathbb{R}\setminus\{0,1\}, \\ 1 - u + u\log u & \text{for } \alpha = 1. \end{cases}$$

Only the parameter $\alpha = \frac{1}{2}$, that is, $g_{1/2}(u) = 2(\sqrt{u} - 1)^2$ provides a genuine distance, namely, the *Hellinger distance*

$$H(Q,P) := \left[\int_{\Omega} \left(\sqrt{q(s)} - \sqrt{p(s)}\right)^2 d\mu(s)\right]^{\frac{1}{2}}.$$

Another important divergence in this class is the *Kullback-Leibler divergence* obtained for $\alpha = 1$ and given by

$$KL(Q,P) := \int_{\Omega} q(s) \log\left[\frac{q(s)}{p(s)}\right] d\mu(s).$$

For other classes of $f-$divergence such as *Matushita's divergence*, *Puri-Vincze divergences* or *Arimoto-type divergences*, see [7], [12] and [11].

Now, for a continuous convex function $f : [0,\infty) \to \mathbb{R}$, consider the $*-$*conjugate* function

$$f^*(u) := uf\left(\frac{1}{u}\right) \quad \text{if } u \in (0,\infty)$$

and

$$f^*(0) := \lim_{u \downarrow 0} f^*(u).$$

It is well known that if f is continuous convex on $[0,\infty)$ then f^* is the same. The following results contain the most basic properties of $f-$divergences (for their proof, we refer to Chapter 1 of [11]).

Theorem 1 (Uniqueness and Symmetry Theorem)**.** *Let f, f_1 be continuous convex functions on $[0,\infty)$. We have*

$$I_{f_1}(Q,P) = I_f(Q,P)$$

for all $P,Q \in \mathcal{P}$ if and only if there exists a constant $c \in \mathbb{R}$ such that $f_1(u) = f(u) + c(u-1)$ for any $u \in [0,\infty)$.

Theorem 2 (Range of Values Theorem)**.** *Let $f : [0,\infty) \to \mathbb{R}$ be a convex function on $[0,\infty)$. For any $P,Q \in \mathcal{P}$ we have the double inequality*

$$f(1) \le I_f(Q,P) \le f(0) + f^*(0). \tag{1.5}$$

(i) *If* $P = Q$, *then the equality holds in the first part of (1.5). If* f *is strictly convex at* 1, *then the equality holds in the first part of (1.5) if and only if* $P = Q$;

(ii) *If* $Q \perp P$, *then the equality holds in the second part of (1.5). If* $f(0) + f^*(0) < \infty$, *then equality holds in the second part of (1.5) if and only if* $Q \perp P$.

For other recent results concerning inequalities for $f-$divergences, see [2] and [4].

2. Inequalities for $2n-$Convex Functions

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space consisting of a set Ω, a $\sigma-$algebra of parts denoted by $\mathcal{A}$ and a countably additive and positive measure μ defined on $\mathcal{A}$ with values in $\mathbb{R} \cup \{\infty\}$.

Assume that the function $w : \Omega \to [0, \infty)$ is $\mu-$measurable with the property that $\int_\Omega w(s)\, d\mu(s) = 1$. We consider the Lebesgue space $L_w(\Omega, \mu) := \{f : \Omega \to \mathbb{R},$ f is $\mu-$measurable and $\int_\Omega w(s)|f(s)|\, d\mu(s) < \infty\}$.

The following result that provides a double Jensen type inequality may be stated.

Theorem 3. *Let* $F : I \subseteq \mathbb{R} \to \mathbb{R}$ *be a* $2n-$*time differentiable function on the interior* $\mathring{I}$ *of the interval* I *and* $n \geq 1$. *Also, assume that* $F^{(2n)}(t) \geq 0$ *for each* $t \in \mathring{I}$, *i.e.,* F *is* $2n-$*convex on* $\mathring{I}$. *If* $f : \Omega \to \mathring{I}$ *is a* $\mu-$*measurable function on* Ω *and such that* f, $\left(f - \int_\Omega w(s) f(s)\, d\mu(s)\right)^k$, $F \circ f$, $F^{(k)} \circ f$ *is in* $L_w(\Omega, \mu)$ *for each* $k = 1, \dots, 2n-1$ *and* $\int_\Omega w(s) f(s)\, d\mu(s) \in \mathring{I}$, *then,*

$$\begin{aligned}
(2.1) \qquad & \sum_{k=1}^{2n-1} \frac{F^{(k)}\left(\int_\Omega w(s) f(s)\, d\mu(s)\right)}{k!} \\
& \qquad \times \int_\Omega w(s) \left[f(s) - \int_\Omega w(z) f(z)\, d\mu(z)\right]^k d\mu(s) \\
& \leq \int_\Omega w(s) F(f(s))\, d\mu(s) - F\left(\int_\Omega w(s) f(s)\, d\mu(s)\right) \\
& \leq \sum_{k=1}^{2n-1} \frac{(-1)^{k+1}}{k!} \\
& \qquad \times \int_\Omega w(s) \left[f(s) - \int_\Omega w(z) f(z)\, d\mu(z)\right]^k F^{(k)}(f(s))\, d\mu(s).
\end{aligned}$$

Proof. We observe that, by Taylor's representation theorem with integral remainder, we have for each $x, a \in \mathring{I}$ that

$$(2.2) \qquad F(x) = \sum_{k=0}^{2n-1} \frac{(x-a)^k}{k!} F^{(k)}(a) + \frac{1}{(2n-1)!} \int_a^x (x-t)^{2n-1} F^{(2n)}(t)\, dt.$$

Since $F^{(2n)}(t) \geq 0$ for any $t \in \mathring{I}$, then on denoting by

$$R(x) := \int_a^x (x-t)^{2n-1} F^{(2n)}(t)\, dt,$$

we observe that $R(x) \geq 0$ for any $x \geq a$. Also, for $x < a$ we can write that

$$R(x) = -\int_x^a (x-t)^{2n-1} F^{(2n)}(t)\, dt = \int_x^a (t-x)^{2n-1} F^{(2n)}(t)\, dt \geq 0$$

showing that in fact $R(x) \geq 0$ for each $x, a \in \mathring{I}$. Therefore, by (2.2) we can state that

$$F(x) \geq F(a) + \sum_{k=1}^{2n-1} \frac{(x-a)^k}{k!} F^{(k)}(a) \tag{2.3}$$

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

Now, if we choose in (2.3) $x = f(s)$, $s \in \Omega$ and $a = \int_\Omega w(z) f(z)\, d\mu(z)$, then we get

$$\begin{aligned} F(f(x)) &\geq F\left(\int_\Omega w(z) f(z)\, d\mu(z)\right) \\ &+ \sum_{k=1}^{2n-1} \frac{\left[f(s) - \int_\Omega w(s) f(s)\, d\mu(s)\right]^k}{k!} \cdot F^{(k)}\left(\int_\Omega w(z) f(z)\, d\mu(z)\right) \end{aligned} \tag{2.4}$$

for each $s \in \Omega$. If we multiply this inequality by $w(s) \geq 0$ and integrate on Ω over the positive measure μ we deduce the first inequality in (2.1).

By changing the place of x with a in (2.3) we also have

$$\sum_{k=1}^{2n-1} \frac{(-1)^{k+1}}{k!} (x-a)^k F^{(k)}(x) + F(a) \geq F(x) \tag{2.5}$$

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

Now, if in (2.5) we choose $x = f(s)$, $s \in \Omega$ and $a = \int_\Omega w(s) f(s)\, d\mu(s)$, we obtain

$$\begin{aligned} &\sum_{k=1}^{2n-1} \frac{(-1)^{k+1}}{k!} \left(f(s) - \int_\Omega w(z) f(z)\, d\mu(z)\right)^k F^{(k)}(f(s)) \\ &\qquad + F\left(\int_\Omega w(z) f(z)\, d\mu(z)\right) \geq F(f)(s) \end{aligned} \tag{2.6}$$

for any $s \in \Omega$.

Finally, if we multiply (2.6) by $w(s) \geq 0$ and integrate on Ω over the positive measure μ, we deduce the second part of the inequality (2.1) and the theorem is proved. ∎

Corollary 1. *If F is twice differentiable and convex and f, $F \circ f$, $F' \circ f \in L_w(\Omega, \mu)$, then*

$$\begin{aligned} 0 &\leq \int_\Omega w(s) F(f(s))\, d\mu(s) - F\left(\int_\Omega w(s) f(s)\, d\mu(s)\right) \\ &\leq \int_\Omega w(s) f(s) F'(f(s))\, d\mu(s) \\ &\qquad - \int_\Omega w(s) f(s)\, d\mu(s) \cdot \int_\Omega w(s) F'(f(s))\, d\mu(s). \end{aligned} \tag{2.7}$$

A similar result has been obtained in [6].

Remark 1. *The discrete case, i.e., where* $\Omega = \{1, \dots, m\}$ *and* μ *is the discrete measure, is of interest and can be stated as:*

$$\begin{aligned} & \sum_{k=1}^{2n-1} \frac{F^{(k)}(\bar{x}_p)}{k!} \cdot \sum_{i=1}^{m} p_i (x_i - \bar{x}_p)^k \\ & \leq \sum_{i=1}^{m} p_i F(x_i) - F(\bar{x}_p) \\ & \leq \sum_{k=1}^{2n-1} \frac{(-1)^{k+1}}{k!} \cdot \sum_{i=1}^{m} p_i (x_i - \bar{x}_p)^k F^{(k)}(x_i), \end{aligned} \tag{2.8}$$

where $x_i \in \mathring{I}$, $p_i \geq 0$ *with* $\sum_{i=1}^{n} p_i = 1$ *and* $\bar{x}_p := \sum_{i=1}^{m} p_i x_i \in \mathring{I}$, *while* $F : I \to \mathbb{R}$ *is as in the statement of Theorem 3. We also notice that if* F *is differentiable and convex, then we can deduce from (2.8) the following reverse of the Jensen inequality:*

$$\begin{aligned} 0 & \leq \sum_{i=1}^{n} p_i f(x_i) - f(\bar{x}_p) \\ & \leq \sum_{i=1}^{n} p_i x_i f'(x_i) - \bar{x}_p \cdot \sum_{i=1}^{n} p_i f'(x_i) \end{aligned} \tag{2.9}$$

that was obtained by Dragomir and Ionescu in 1994, see [5].

Remark 2. *We recall that a function* $f : (a, b) \to \mathbb{R}$ *which has derivatives of all orders is said to be absolutely monotonic if* $f^{(n)}(t) \geq 0$ *for all* $t \in (a, b)$ *and* $n = 0, 1, \dots,$ *and* f *is called completely monotonic if* $(-1)^n f^{(n)}(t) \geq 0$ *for all* $t \in (a, b)$ *and* $n = 0, 1, 2, \dots$. *It is therefore obvious that Theorem 3 can be applied for any absolutely monotonic or completely monotonic function* $F : I \to \mathbb{R}$ *and any* $n \geq 1$. *However, the class of functions* F *for which Theorem 3 is valid is much larger. One can choose for instance* $2n-$*differentiable functions* $g : (a, b) \to \mathbb{R}$ *with the property that* $m := \inf_{t \in (a,b)} g^{(2n)}(t) > -\infty$ *and consider the new function* $F : (a, b) \to \mathbb{R}$, $F(t) := g(t) - \frac{m \cdot t^{2n}}{(2n)!}$, *which is* $2n-$*differentiable and* $F^{(2n)}(t) = g^{(2n)}(t) - m \geq 0$ *for any* $t \in (a, b)$.

Remark 3. *The discrete inequality (2.8) can be utilized to provide various inequalities for means.*

For instance, if we choose $F(t) = -\log t$, *then*

$$F^{(k)}(t) = \frac{(-1)^k (k-1)!}{t^k}, \quad k \geq 1, \ t > 0$$

and then for any $x_i, p_i > 0$, $i \in \{1, \dots, n\}$ *with* $\sum_{i=1}^{n} p_i = 1$ *and* $\bar{x}_p := \sum_{i=1}^{n} p_i x_i$ *and* $G(p; x) := \prod_{i=1}^{n} x_i^{p_i}$, *we have:*

$$\begin{aligned} \sum_{k=1}^{2n-1} \frac{(-1)^k}{k} \sum_{i=1}^{m} p_i \left(\frac{x_i}{\bar{x}_p} - 1 \right)^k & \leq \log \left[\frac{\bar{x}_p}{G(p; x)} \right] \\ & \leq \sum_{k=1}^{2n-1} \frac{(-1)^{k+1}}{k} \sum_{i=1}^{m} p_i \left(\frac{\bar{x}_p}{x_i} - 1 \right)^k, \end{aligned} \tag{2.10}$$

for any $n \geq 1$.

We remark that for $n=1$, we obtained the simpler inequality:

$$0 \le \log\left[\frac{\bar{x}_p}{G(p;x)}\right] \le \bar{x}_p \cdot \bar{h}_p - 1, \tag{2.11}$$

where $\bar{h}_p := \sum_{i=1}^{m} \frac{p_i}{x_i}$ is the harmonic mean of x_i with the weights p_i. This is a known result, see for instance [5].

For $F(t) = \exp(t)$, $t \in \mathbb{R}$, we get from (2.8) the following result as well:

$$\begin{aligned} &\exp(\bar{x}_p) \sum_{k=1}^{2n-1} \frac{(-1)^k}{k!} \sum_{i=1}^{m} p_i (x_i - \bar{x}_p)^k \\ &\le \sum_{i=1}^{m} p_i \exp(x_i) - \exp(\bar{x}_p) \\ &\le \sum_{k=1}^{2n-1} \frac{(-1)^{k+1}}{k!} \sum_{i=1}^{m} p_i (x_i - \bar{x}_p) \exp(x_i). \end{aligned} \tag{2.12}$$

For $n=1$, the inequality (2.12) produces the following particular case of interest:

$$\begin{aligned} 0 &\le \sum_{i=1}^{m} p_i \exp(x_i) - \exp(\bar{x}_p) \\ &= \sum_{i=1}^{m} p_i x_i \exp(x_i) - x \cdot \sum_{i=1}^{m} p_i \exp(x_i). \end{aligned}$$

One can state other particular inequalities by choosing elementary function for which $F^{(2n)}(t) \ge 0$ on the given interval. The details are omitted.

3. Inequalities for $F-$Divergences

We consider now a function $F : [0,\infty) \to \mathbb{R}$ which has the $2n-$derivative $F^{(2n)}$ nonnegative on $(0,\infty)$. If P, Q are two probability distributions as in the introduction, we can define the following divergences:

$$I_{\chi^k}(Q,P) := \int_\Omega p(s)\left(\frac{q(s)}{p(s)} - 1\right)^k d\mu(s) \tag{3.1}$$

and

$$I_{g_k}(Q,P) := \int_\Omega p(s)\, g_k\left(\frac{q(s)}{p(s)}\right) d\mu(s), \tag{3.2}$$

where the function g_k is defined by

$$g_k(t) := (t-1)^k F^{(k)}(t), \quad t \in [0,\infty),\ k \ge 1 \tag{3.3}$$

and is arguably simpler than the function F which generates it. This indeed happens if $F(t) = -\log t$ since the derivatives $F^{(k)}(t)$ are in this case rational functions. The same applies if $F(t) = \int_0^t u(\tau)\, d\tau$ and the integral cannot be represented by elementary functions.

The following result provides upper and lower bounds for the $f-$divergence

$$I_F(Q,P) := \int_\Omega p(s)\, F\left(\frac{q(s)}{p(s)}\right) d\mu(s)$$

in terms of the divergences introduced in equations (3.1) and (3.3) above.

Theorem 4. *Let $F:[0,\infty)\to\mathbb{R}$ be a $2n-$differentiable function, $n\geq 1$ and such that $F^{(2n)}(t)\geq 0$ on $(0,\infty)$. Then*

$$\sum_{k=1}^{2n-1}\frac{F^{(k)}(1)}{k!}I_{\chi^k}(Q,P)\leq I_F(Q,P)-F(1) \tag{3.4}$$
$$\leq\sum_{k=1}^{2n-1}\frac{(-1)^{k+1}}{k!}I_{g_k}(Q,P).$$

Proof. We apply Theorem 3 for the choices $w(s)=p(s)$, $f(x)=\frac{q(s)}{p(s)}$, $s\in\Omega$ to get:

$$\sum_{k=1}^{2n-1}\frac{F^{(k)}\left(\int_\Omega q(s)\,d\mu(s)\right)}{k!}\cdot\int_\Omega p(s)\left[\frac{q(s)}{p(s)}-\int_\Omega q(z)\,d\mu(z)\right]^k d\mu(s) \tag{3.5}$$
$$\leq\int_\Omega p(s)F\left(\frac{q(s)}{p(s)}\right)d\mu(s)-F\left(\int_\Omega q(s)\,d\mu(s)\right)$$
$$\leq\sum_{k=1}^{2n-1}\frac{(-1)^{k+1}}{k!}\cdot\int_\Omega p(s)\left[\frac{q(s)}{p(s)}-\int_\Omega q(z)\,d\mu(z)\right]^k F^{(k)}\left(\frac{q(s)}{p(s)}\right)d\mu(s)$$

and since $\int_\Omega q(s)\,d\mu(s)=1$, hence, with notations (3.1) and (3.3), we observe that (3.5) is exactly the desired inequality (3.4). ∎

We observe that for $n=1$, $I_{\chi^1}(Q,P)=0$ and

$$I_{g_1}(Q,P)=\int_\Omega p(s)\left(\frac{q(s)}{p(s)}-1\right)F'\left(\frac{q(s)}{p(s)}\right)d\mu(s)$$
$$=\int_\Omega (q(s)-p(s))\,F'\left(\frac{q(s)}{p(s)}\right)d\mu(s),$$

therefore the following particular case may be stated:

Corollary 2. *Assume that $F:[0,\infty)\to\mathbb{R}$ is continuous and twice differentiable on $[0,\infty)$. If F is convex on $[0,\infty)$, then the following inequality can be stated:*

$$0\leq I_F(Q,P)-F(1)\leq\delta_{F'}(Q,P), \tag{3.6}$$

where

$$\delta_{F'}(Q,P):=\int_\Omega (q(s)-p(s))\,F'\left(\frac{q(s)}{p(s)}\right)d\mu(s)$$

is the general $\delta-$divergence measure introduced in the recent paper [4] *by the first author.*

For the definition of $\delta-$divergence measures and some of its fundamental properties, see [4].

Remark 4. *It is well known that the Gamma function*

$$\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt,\quad x>0$$

is logarithmic-convex on $(0,\infty)$. Therefore, one can consider the divergences generated by the convex function $\log\Gamma$, *i.e.,*

$$I_{\log\Gamma}(Q,P):=\int_\Omega p(s)\log\Gamma\left(\frac{q(s)}{p(s)}\right)d\mu(s).$$

If we denote by $\Psi(t) = \frac{d}{dt}[\log\Gamma(t)] = \frac{\Gamma'(t)}{\Gamma(t)}$, which is well known in the literature as the "psi" or "digamma function", then utilizing Corollary 2, we can conveniently connect the $f-$divergence and $\log\Gamma$ with the $\delta-$divergence of Ψ via an inequality. We have therefore the inequality:

$$0 \le I_{\log\Gamma}(Q,P) \le \delta_{\Psi}(Q,P), \tag{3.7}$$

for any $P, Q \in \mathcal{P}$, where, as above

$$\delta_{\Psi}(Q,P) := \int_{\Omega} (q(s) - p(s)) \Psi\left(\frac{q(s)}{p(s)}\right) d\mu(s).$$

If we consider now the *zeta function*

$$\zeta(x) := \sum_{n=0}^{\infty} \frac{1}{n^x}, \quad x > 1$$

then we have that $\zeta(\cdot + 1)$ is log-convex and if we denote by

$$m(t) := \frac{d}{dt}[\log\zeta(t+1)] = \frac{\zeta'(t+1)}{\zeta(t+1)}, \quad t > 0,$$

then the following representation is well known

$$m(t) = -\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^{t+1}}, \quad t > 1,$$

where $\Lambda(n)$ is the *van Mongoldt function*, i.e.,

$$\Lambda(n) := \begin{cases} \log p & \text{if } n = p^k \quad (p \text{ prime}, k \ge 1), \\ 0 & \text{otherwise.} \end{cases}$$

We can label $m(t)$ to be a "distance function" by following the same approach as that for gamma.

We can then introduce the $f-$divergence of $\log\zeta(\cdot + 1)$ and state, by utilizing Corollary 2 that

$$0 \le I_{\log\zeta(\cdot+1)}(Q,P) \le \delta_m(Q,P), \tag{3.8}$$

for any $P, Q \in \mathcal{P}$, where $\delta_m(Q,P)$ is the $\delta-$divergence of m defined above.

Similar results may be considered for other log-convex functions for which the log-derivatives are function of specific interest and generating $\delta-$divergences which can be easier to calculate/estimate. In the next section we give a different type of applications of $F-$divergence inequalities.

4. Applications to product measures

In this section we prove the following

Proposition 1. *Let a_n and b_n be sequences of real numbers such that $|a_n| < 1$, $|b_n| < 1$. Then for all positive integers n we have*

$$\prod_{k=1}^{n} \left(\frac{1+b_k}{1+a_k}\right)^{\frac{1+a_k}{2}} \left(\frac{1-b_k}{1-a_k}\right)^{\frac{1-a_k}{2}} \ge \exp\left[1 - \prod_{k=1}^{n}\left(1 + \frac{(a_k-b_k)^2}{1-b_k^2}\right)\right]. \tag{4.1}$$

Before proving (4.1) let us state an application of it.

Corollary 3. *Suppose that the sequences a_n and b_n are as above. If*

$$\sum_{n=1}^{\infty} \frac{(a_n - b_n)^2}{1 - b_n^2} < \infty, \tag{4.2}$$

then the infinite product

$$\prod_{n=1}^{\infty} \Big(\frac{1 + b_n}{1 + a_n}\Big)^{\frac{1+a_n}{2}} \Big(\frac{1 - b_n}{1 - a_n}\Big)^{\frac{1-a_n}{2}}, \tag{4.3}$$

converges.

Proof. It is clear that

$$0 < \Big(\frac{1 + b_k}{1 + a_k}\Big)^{\frac{1+a_k}{2}} \Big(\frac{1 - b_k}{1 - a_k}\Big)^{\frac{1-a_k}{2}} < 1.$$

Then use condition (4.2) in combination with (4.1) to complete the proof. ∎

Remark 5. *We observe that the conditions*

$$\sum_{n=1}^{\infty} (a_n - b_n)^2 < \infty \quad \textit{and} \quad \limsup |b_n| < 1 \tag{4.4}$$

imply (4.2).

We now give a proof of (4.1)

Proof. Let $r_n(x),\ n = 1, 2, \ldots$ be the Rademacher functions defined on the interval [0, 1] by the relation $r_n(x) = \operatorname{sign} \sin 2^n \pi x$. These functions form an orthonormal set in the Hilbert space $L^2([0,\ 1])$ and also

$$\int_0^1 r_n(x) dx = 0, \quad n = 1, 2, \ldots.$$

The functions $r_n(x)$ are also independent random variables in the probability measure space $([0,\ 1],\ \mathcal{B},\ \lambda)$, where λ is the Lebesgue measure on the σ-algebra $\mathcal{B}$ of Borel subsets of [0, 1]. Let a_n and b_n be sequences of real numbers such that $|a_n| < 1$, $|b_n| < 1$. For every $n \in \mathbb{N}$ the relations

$$dP = \prod_{k=1}^{n} (1 + a_k\, r_k)\, d\lambda, \quad dQ = \prod_{k=1}^{n} (1 + b_k\, r_k)\, d\lambda \quad , \tag{4.5}$$

define probability measures on $([0,\ 1],\ \mathcal{B},\ \lambda)$, which are absolutely continuous with respect to λ.

For $F(t) = -\log t$ we will calculate the F-divergence of the probability distributions P and Q defined in (4.5), *viz.*

$$\begin{aligned} I_F(Q, P) &= -\int_0^1 \prod_{k=1}^{n} (1 + a_k\, r_k(x)) \log\Big(\frac{\prod_{k=1}^n (1 + b_k\, r_k(x))}{\prod_{k=1}^n (1 + a_k\, r_k(x))}\Big) dx \\ &= -\int_0^1 \log\Big(\frac{\prod_{k=1}^n (1 + b_k\, r_k(x))}{\prod_{k=1}^n (1 + a_k\, r_k(x))}\Big) dP\,. \end{aligned} \tag{4.6}$$

We observe that

$$\log(1 + b_k\, r_k(x)) = \frac{1}{2} \log(1 - b_k^2) + \frac{1}{2} \log\Big(\frac{1 + b_k}{1 - b_k}\Big) r_k(x)\,. \tag{4.7}$$

It is easy to see that

$$\int_0^1 r_k(x)\, dP = a_k, \quad k = 1, 2, \ldots$$

Then by (4.7) we obtain

$$\int_0^1 \sum_{k=1}^n \log(1 + b_k\, r_k(x))\, dP = \frac{1}{2} \sum_{k=1}^n \log(1 - b_k^2) + \frac{1}{2} \sum_{k=1}^n a_k \, \log \frac{1 + b_k}{1 - b_k} \tag{4.8}$$

Hence, by (4.6) we find that

$$-I_F(Q, P) = \frac{1}{2} \sum_{k=1}^n \log \frac{1 - b_k^2}{1 - a_k^2} + \frac{1}{2} \sum_{k=1}^n a_k \, \log \frac{(1 + b_k)(1 - a_k)}{(1 - b_k)(1 + a_k)},$$

and finally

$$I_F(Q, P) = -\log \prod_{k=1}^n \Big(\frac{1 + b_k}{1 + a_k}\Big)^{\frac{1+a_k}{2}} \Big(\frac{1 - b_k}{1 - a_k}\Big)^{\frac{1-a_k}{2}}. \tag{4.9}$$

Now we set

$$p(x) = \prod_{k=1}^n (1 + a_k\, r_k(x)), \quad q(x) = \prod_{k=1}^n (1 + b_k\, r_k(x))$$

and calculate the δ-divergence measure of P and Q. We have

$$\begin{aligned} \delta_{F'}(Q, P) &= -\int_0^1 (q(x) - p(x))\, \frac{p(x)}{q(x)}\, dx = -1 + \int_0^1 \frac{p(x)^2}{q(x)}\, dx \\ &= -1 + \int_0^1 \prod_{k=1}^n \frac{(1 + a_k\, r_k(x))^2\, (1 - b_k r_k(x)}{1 - b_k^2}\, dx \\ &= -1 + \prod_{k=1}^n \Big(1 + \frac{(a_k - b_k)^2}{1 - b_k^2}\Big). \end{aligned} \tag{4.10}$$

Combining (4.9) with (4.10) and using the inequality (3.6) we conclude the proof of (4.1). ∎

5. Remarks

(1) Suppose that $b_k = 0$ for all $k \in \mathbb{N}$. Then inequality (4.1) is equivalent to

$$\prod_{k=1}^n (1 + a_k^2) \geq 1 + \frac{1}{2} \log \prod_{k=1}^n (1 + a_k)^{1 + a_k}\, (1 - a_k)^{1 - a_k}. \tag{5.1}$$

Inequality (5.1) has the following interpretation:

Let μ be the Borel probability measure on [0, 1] defined by

$$d\mu = \lim_{n \to \infty} \prod_{k=1}^n (1 + a_k\, r_k(x))\, dx, \tag{5.2}$$

where the limit is in the weak$\star$ sense. This measure is absolutely continuous with respect to the Lebesgue measure if and only if $\sum_{n=1}^\infty a_n^2 < \infty$. Compare the paper [8].

Suppose now that M is a Borel subset of the support of μ such that $\mu(M) > 0$. It follows from the results of [9] (see also [1]) that for the Hausdorff dimension of M, $\dim M$, one has

$$\dim M = 1 - \limsup \frac{1}{n \log 4} \log \prod_{k=1}^{n} (1 + a_k)^{1+a_k} (1 - a_k)^{1-a_k} . \tag{5.3}$$

It follows from this and (5.1) that in the case where $\sum_{n=1}^{\infty} a_n^2 < \infty$ for every Borel subset M of the support of μ with $\mu(M) > 0$, we have $\dim M = 1$. Of course, this happens because in this case μ is absolutely continuous with respect to Lebesgue measure λ and thus $\lambda(M) > 0$.

(2) Let the probability measure μ be defined by (5.2) and ν be a product measure of the same type, that is

$$d\nu = \lim_{n\to\infty} \prod_{k=1}^{n} (1 + b_k\, r_k(x))\, dx\, . \tag{5.4}$$

In the case where $(a_n),\ (b_n) \in \ell_2$ we can calculate the F-divergence measure of the probability distributions $\mu,\ \nu$ for $F(t) = -\log t$. Indeed, a small adaptation of the proof of Proposition 1 yields

$$I_F(\nu,\, \mu) = -\log \prod_{k=1}^{\infty} \Big(\frac{1 + b_k}{1 + a_k}\Big)^{\frac{1+a_k}{2}} \Big(\frac{1 - b_k}{1 - a_k}\Big)^{\frac{1-a_k}{2}} . \tag{5.5}$$

Since the sequences $(a_n),\ (b_n) \in \ell_2$, both μ and ν is absolutely continuous with respect to the Lebesgue measure λ and the condition (4.2) is satisfied therefore by Corollary 3, the infinite product above converges.

Finally, we note that we can obtain results analogous to (4.1), (5.1) and (5.5) using the more general product measures considered in [9] and [10].

References

[1] A. BISBAS and C. KARANIKAS, On the Hausdorff dimension of Rademacher-Riesz products. *Monatsh. Math.* **110**, (1990), 15–21.

[2] P. CERONE, S.S. DRAGOMIR and F. ÖSTERREICHER, Bounds of extended $f-$divergences for a variety of classes, *Kybernetika* (Prague), **40**(6) (2004), 745-756.

[3] I. CSISZÁR, Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. (German) *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **8** 1963 85–108.

[4] S.S. DRAGOMIR, Some general divergence measures for probability distributions, *Acta Math. Hung.*, **109**(**4**) (2005), 331-345.

[5] S.S. DRAGOMIR and N.M. IONESCU, Some converse of Jensen's inequality and applications, *Rev. Anal. Numér. Théor. Approx.*, **22**(1) (1994), 71-78.

[6] S. S. DRAGOMIR and C. J. GOH, Some counterpart inequalities for a functional associated with Jensen's inequality, *J. Ineq. Appl.* **1** (1997), 311–325.

[7] P. KAFKA, F. ÖSTERREICHER and I. VINCZE, On powers of f-divergences defining a distance. *Studia Sci. Math. Hungar.* **26** (1991), no. 4, 415–422.

[8] C. KARANIKAS and S. KOUMANDOS, Continuous singular measures with absolutely continuous convolution squares on locally compact groups. *Illinois J. Math.* **35** (1991), no. 3, 490–495.

[9] C. KARANIKAS and S. KOUMANDOS, On a generalized entropy's formula. *Results in Mathematics* **18** (1990), no. 3-4, 254–263.

[10] S. KOUMANDOS, L^p-Convergence of a certain class of product martingales. *Proc. Edinburgh Math. Soc* **37** (1994), no. 2, 243–254.

[11] F. LIESE and I. VAJDA, *Convex Statistical Distances.* With German, Teubner-Texte zur Mathematik [Teubner Texts in Mathematics], 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987. 224 pp.

[12] F. ÖSTERREICHER and I. VAJDA, A new class of metric divergences on probability spaces and its applicability in statistics. *Ann. Inst. Statist. Math.* **55** (2003), no. 3, 639–653.

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

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

URL: http://rgmia.vu.edu.au/dragomir

Department of Mathematics and Statistics, The University of Cyprus, P.O. Box. 20537, 1678 Nicosia, CYPRUS

E-mail address: skoumand@ucy.ac.cy

URL: http://www.mas.ucy.ac.cy/teachers/koumantos.htm