Superadditivity of the Jensen integral inequality with applications

S. S. Dragomir

SUPERADDITIVITY OF THE JENSEN INTEGRAL INEQUALITY WITH APPLICATIONS

S. S. DRAGOMIR

Abstract. The superadditivity and monotonicity of two functionals associated to the celebrated Jensen's integral inequality for convex functions with applications for Hölder's inequality and f-divergence measures in information theory are given.

1. INTRODUCTION

Let $(\Omega, \mathcal{A}, \mu)$ be a measurable space consisting of a set Ω, a σ–algebra $\mathcal{A}$ of parts of Ω and a countably additive and positive measure μ on $\mathcal{A}$ with values in $\mathbb{R} \cup \{\infty\}$. For a μ−measurable function $w : \Omega \to \mathbb{R}$, with $w(x) \geq 0$ for μ–a.e. (almost every) $x \in \Omega$, consider the Lebesgue space

$$L_w(\Omega, \mu) := \{f : \Omega \to \mathbb{R}, \ f \text{ is } \mu\text{-measurable and } \int_\Omega w(x)|f(x)| \, d\mu(x) < \infty\}.$$

For simplicity of notation we write everywhere in the sequel $\int_\Omega w \, d\mu$ instead of $\int_\Omega w(x) \, d\mu(x)$.

In order to provide a reverse of the celebrated Jensen's integral inequality for convex functions, S.S. Dragomir obtained in 2002 [10] the following result:

Theorem 1. *Let $\Phi : [m, M] \subset \mathbb{R} \to \mathbb{R}$ be a differentiable convex function on (m, M) and $f : \Omega \to [m, M]$ so that $\Phi \circ f, \ f, \ \Phi' \circ f, \ (\Phi' \circ f) f \in L_w(\Omega, \mu)$, where $w \geq 0$ μ-a.e. (almost everywhere) on Ω with $\int_\Omega w \, d\mu = 1$. Then we have the inequality:*

$$
\begin{aligned}
0 &\leq \int_\Omega w(\Phi \circ f) \, d\mu - \Phi\left(\int_\Omega w f \, d\mu\right) \\
&\leq \int_\Omega w(\Phi' \circ f) f \, d\mu - \int_\Omega w(\Phi' \circ f) \, d\mu \int_\Omega w f \, d\mu
\end{aligned}
\tag{1.1}
$$

$$\leq \frac{1}{2} \left[\Phi'(M) - \Phi'(m) \right] \int_\Omega w \left| f - \int_\Omega w f d\mu \right| d\mu.$$

For a generalization of the first inequality in (1.1) without the differentiability assumption and the derivative Φ' replaced with a selection φ from the subdifferential $\partial \Phi$, see the paper [24] by C.P. Niculescu.

Remark 1. We notice that in the discrete case, the inequality between the first and the second term in (1.1) was proved in 1994 by Dragomir & Ionescu, see [12].

Utilising a different approach than in [10], we obtained in [11] the following two results that provide other upper bounds for the Jensen's difference:

$$\int_\Omega w (\Phi \circ f) d\mu - \Phi \left(\int_\Omega w f d\mu \right).$$

Theorem 2. *Let $\Phi : I \to \mathbb{R}$ be a continuous convex function on the interval of real numbers I and $m, M \in \mathbb{R}$, $m < M$ with $[m, M] \subset \mathring{I}$, $\mathring{I}$ is the interior of I. If $f : \Omega \to \mathbb{R}$ is μ-measurable, satisfies the bounds*

$$-\infty < m \leq f(x) \leq M < \infty \text{ for } \mu\text{-a.e. } x \in \Omega$$

and is such that $f, \Phi \circ f \in L_w (\Omega, \mu)$, then

$$0 \leq \int_\Omega w (\Phi \circ f) d\mu - \Phi \left(\bar{f}_{\Omega,w} \right) \tag{1.2}$$

$$\leq \frac{\left(M - \bar{f}_{\Omega,w} \right) \left(\bar{f}_{\Omega,w} - m \right)}{M - m} \sup_{t \in (m,M)} \Psi_\Phi (t; m, M)$$

$$\leq \left(M - \bar{f}_{\Omega,w} \right) \left(\bar{f}_{\Omega,w} - m \right) \frac{\Phi'_- (M) - \Phi'_+ (m)}{M - m}$$

$$\leq \frac{1}{4} (M - m) \left[\Phi'_- (M) - \Phi'_+ (m) \right],$$

where $\bar{f}_{\Omega,w} := \int_\Omega w(x) f(x) d\mu(x) \in [m, M]$ and $\Psi_\Phi (\cdot; m, M) : (m, M) \to \mathbb{R}$ is defined by

$$\Psi_\Phi (t; m, M) = \frac{\Phi(M) - \Phi(t)}{M - t} - \frac{\Phi(t) - \Phi(m)}{t - m}.$$

We also have the inequality

$$0 \leq \int_\Omega w (\Phi \circ f) d\mu - \Phi \left(\bar{f}_{\Omega,w} \right) \leq \frac{1}{4} (M - m) \Psi_\Phi \left(\bar{f}_{\Omega,w}; m, M \right) \tag{1.3}$$

$$\leq \frac{1}{4} (M - m) \left[\Phi'_- (M) - \Phi'_+ (m) \right],$$

provided that $\bar{f}_{\Omega,w} \in (m, M)$.

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

$$0 \leq \int_{\Omega} w \left(\Phi \circ f\right) d\mu \left(x\right) - \Phi \left(\bar{f}_{\Omega,w}\right) \tag{1.4}$$

$$\leq 2 \max \left\{ \frac{M - \bar{f}_{\Omega,w}}{M - m}, \frac{\bar{f}_{\Omega,w} - m}{M - m} \right\} \left[\frac{\Phi \left(m\right) + \Phi \left(M\right)}{2} - \Phi \left(\frac{m + M}{2}\right) \right]$$

$$\leq \frac{1}{2} \max \left\{ M - \bar{f}_{\Omega,w}, \bar{f}_{\Omega,w} - m \right\} \left[\Phi'_{-} \left(M\right) - \Phi'_{+} \left(m\right) \right].$$

Motivated by these results we establish in the current paper some refinements and reverses of the Jensen integral inequality by capitalizing on the superadditivity and monotonicity properties of two associated functionals. Application in connection with the Hölder inequality and for f-divergence measures in Information Theory are provided as well.

2. Superadditivity and Monotonicity Properties

For a μ-measurable function $w : \Omega \to \mathbb{R}$, with $w\left(x\right) \geq 0$ for μ-a.e. $x \in \Omega$ and $\int_{\Omega} w d\mu > 0$ we consider the functional

$$J \left(w; \Phi, f\right) := \int_{\Omega} w \left(\Phi \circ f\right) d\mu - \Phi \left(\frac{\int_{\Omega} w f d\mu}{\int_{\Omega} w d\mu}\right) \int_{\Omega} w d\mu \geq 0, \tag{2.1}$$

where $\Phi : I \to \mathbb{R}$ is a continuous convex function on the interval of real numbers I, $f : \Omega \to \mathbb{R}$ is μ-measurable and such that $f, \Phi \circ f \in L_w \left(\Omega, \mu\right)$.

Theorem 4. *Let $w_i : \Omega \to \mathbb{R}$, with $w_i \left(x\right) \geq 0$ for $\mu - a.e.$ (almost every) $x \in \Omega$ and $\int_{\Omega} w_i d\mu > 0, i \in \{1, 2\}$. If $\Phi : I \to \mathbb{R}$ is a continuous convex function on the interval of real numbers I, $f : \Omega \to \mathbb{R}$ is μ-measurable and such that $f, \Phi \circ f \in L_{w_1} \left(\Omega, \mu\right) \cap L_{w_2} \left(\Omega, \mu\right)$, then*

$$J \left(w_1 + w_2; \Phi, f\right) \geq J \left(w_1; \Phi, f\right) + J \left(w_2; \Phi, f\right) \geq 0 \tag{2.2}$$

i.e., J is a superadditive functional of weights.

Moreover, if $w_2 \geq w_1 \geq 0$ $\mu - a.e.$ on Ω, then

$$J \left(w_2; \Phi, f\right) \geq J \left(w_1; \Phi, f\right) \geq 0, \tag{2.3}$$

i.e., J is a monotonic nondecreasing functional of weights.

Proof. Utilising the convexity property of Φ we have successively

$$J \left(w_1 + w_2; \Phi, f\right) \tag{2.4}$$

$$= \int_{\Omega} \left(w_1 + w_2\right) \left(\Phi \circ f\right) d\mu - \Phi \left(\frac{\int_{\Omega} \left(w_1 + w_2\right) f d\mu}{\int_{\Omega} \left(w_1 + w_2\right) d\mu}\right) \int_{\Omega} \left(w_1 + w_2\right) d\mu$$

$$= \int_{\Omega} w_1 \left(\Phi \circ f\right) d\mu + \int_{\Omega} w_2 \left(\Phi \circ f\right) d\mu$$

$$-\Phi\left(\frac{\int_\Omega w_1 d\mu \cdot \frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu} + \int_\Omega w_2 d\mu \cdot \frac{\int_\Omega w_2 f d\mu}{\int_\Omega w_2 d\mu}}{\int_\Omega (w_1 + w_2) d\mu}\right) \int_\Omega (w_1 + w_2) d\mu$$

$$\geq \int_\Omega w_1 (\Phi \circ f) d\mu + \int_\Omega w_2 (\Phi \circ f) d\mu$$

$$- \left[\frac{\int_\Omega w_1 d\mu}{\int_\Omega (w_1 + w_2) d\mu} \Phi\left(\frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu}\right) + \frac{\int_\Omega w_2 d\mu}{\int_\Omega (w_1 + w_2) d\mu} \Phi\left(\frac{\int_\Omega w_2 f d\mu}{\int_\Omega w_2 d\mu}\right)\right]$$

$$\times \int_\Omega (w_1 + w_2) d\mu$$

$$= \int_\Omega w_1 (\Phi \circ f) d\mu - \Phi\left(\frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu}\right) \int_\Omega w_1 d\mu$$

$$+ \int_\Omega w_2 (\Phi \circ f) d\mu - \Phi\left(\frac{\int_\Omega w_2 f d\mu}{\int_\Omega w_2 d\mu}\right) \int_\Omega w_2 d\mu$$

$$= J(w_1; \Phi, f) + J(w_2; \Phi, f)$$

which proves the superadditivity property.

Now, if $w_2 \geq w_1 \geq 0$, then on applying the superadditivity property we have

$$J(w_2; \Phi, f) = J(w_1 + (w_2 - w_1); \Phi, f) \geq J(w_1; \Phi, f) + J(w_2 - w_1; \Phi, f)$$
$$\geq J(w_1; \Phi, f)$$

since by the Jensen's inequality for the positive weights we have $J(w_2 - w_1; \Phi, f) \geq 0$. $\qquad\square$

The above theorem has a simple however interesting consequence that provides both a refinement and a reverse for the Jensen's integral inequality:

Corollary 1. *Let $w_i : \Omega \to \mathbb{R}$, with $w_i(x) \geq 0$ for $\mu - a.e.$ $x \in \Omega$, $\int_\Omega w_i d\mu > 0, i \in \{1, 2\}$ and there exists the nonnegative constants γ, Γ such that*

$$0 \leq \gamma \leq \frac{w_2}{w_1} \leq \Gamma < \infty \; \mu\text{-a.e. on } \Omega. \tag{2.5}$$

If $\Phi : I \to \mathbb{R}$ is a continuous convex function on the interval of real numbers I, $f : \Omega \to \mathbb{R}$ is μ-measurable and such that $f, \Phi \circ f \in L_{w_1}(\Omega, \mu) \cap L_{w_2}(\Omega, \mu)$, then

$$0 \leq \gamma \cdot \left[\int_\Omega w_1 (\Phi \circ f) d\mu - \Phi\left(\frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu}\right) \int_\Omega w_1 d\mu\right] \tag{2.6}$$

$$\leq \int_\Omega w_2 (\Phi \circ f) d\mu - \Phi\left(\frac{\int_\Omega w_2 f d\mu}{\int_\Omega w_2 d\mu}\right) \int_\Omega w_2 d\mu$$

$$\leq \Gamma \cdot \left[\int_\Omega w_1 (\Phi \circ f) d\mu - \Phi\left(\frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu}\right) \int_\Omega w_1 d\mu\right]$$

or, equivalently,

$$0 \le \gamma \cdot \frac{\int_\Omega w_1 d\mu}{\int_\Omega w_2 d\mu} \left[\frac{\int_\Omega w_1 (\Phi \circ f) d\mu}{\int_\Omega w_1 d\mu} - \Phi \left(\frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu} \right) \right] \qquad (2.7)$$

$$\le \frac{\int_\Omega w_2 (\Phi \circ f) d\mu}{\int_\Omega w_2 d\mu} - \Phi \left(\frac{\int_\Omega w_2 f d\mu}{\int_\Omega w_2 d\mu} \right)$$

$$\le \Gamma \cdot \frac{\int_\Omega w_1 d\mu}{\int_\Omega w_2 d\mu} \left[\frac{\int_\Omega w_1 (\Phi \circ f) d\mu}{\int_\Omega w_1 d\mu} - \Phi \left(\frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu} \right) \right].$$

Proof. From (2.5) we have $\gamma w_1 \le w_2 \le \Gamma w_1 < \infty$ μ-a.e. on Ω and by the monotonicity property (2.3) we get

$$J (\Gamma w_1; \Phi, f) \ge J (w_2; \Phi, f) \ge J (\gamma w_1; \Phi, f). \qquad (2.8)$$

Since the the functional is positive homogeneous, namely $J (\alpha w; \Phi, f) = \alpha J (w; \Phi, f)$, then we get from (2.8) the desired result (2.6). □

Remark 2. Assume that $\mu (\Omega) < \infty$ and let $w : \Omega \to \mathbb{R}$, with $w (x) \ge 0$ for $\mu -$ a.e. $x \in \Omega$, $\int_\Omega w d\mu > 0$ and w is essentially bounded, i.e. ess $\inf_{x \in \Omega} w (x)$ and ess $\sup_{x \in \Omega} w (x)$ are finite.

If $\Phi : I \to \mathbb{R}$ is a continuous convex function on the interval of real numbers I, $f : \Omega \to \mathbb{R}$ is μ-measurable and such that $f, \Phi \circ f \in L_w (\Omega, \mu) \cap L (\Omega, \mu)$, then

$$0 \le \frac{\text{ess } \inf_{x \in \Omega} w (x)}{\frac{1}{\mu(\Omega)} \int_\Omega w d\mu} \left[\frac{\int_\Omega (\Phi \circ f) d\mu}{\mu (\Omega)} - \Phi \left(\frac{\int_\Omega f d\mu}{\mu (\Omega)} \right) \right] \qquad (2.9)$$

$$\le \frac{\int_\Omega w (\Phi \circ f) d\mu}{\int_\Omega w d\mu} - \Phi \left(\frac{\int_\Omega w f d\mu}{\int_\Omega w d\mu} \right)$$

$$\le \frac{\text{ess } \sup_{x \in \Omega} w (x)}{\frac{1}{\mu(\Omega)} \int_\Omega w d\mu} \left[\frac{\int_\Omega (\Phi \circ f) d\mu}{\mu (\Omega)} - \Phi \left(\frac{\int_\Omega f d\mu}{\mu (\Omega)} \right) \right].$$

This result can be used to provide the following result related to the Hermite-Hadamard inequality for convex functions that states that

$$\frac{1}{b - a} \int_a^b \Phi (t) dt \ge \Phi \left(\frac{a + b}{2} \right)$$

for any convex function $\Phi : [a, b] \to \mathbb{R}$.

Indeed , if $w : [a, b] \to [0, \infty)$ is Lebesgue integrable, then we have

$$0 \le \frac{\text{ess } \inf_{x \in [a,b]} w (x)}{\frac{1}{b-a} \int_a^b w (t) dt} \left[\frac{1}{b - a} \int_a^b \Phi (t) dt - \Phi \left(\frac{a + b}{2} \right) \right] \qquad (2.10)$$

$$\le \frac{\int_a^b w (t) \Phi (t) dt}{\int_a^b w (t) dt} - \Phi \left(\frac{\int_\Omega w (t) t dt}{\int_a^b w (t) dt} \right)$$

$$\leq \frac{\operatorname{ess\,sup}_{x \in [a,b]} w(x)}{\frac{1}{b-a} \int_a^b w(t)\, dt} \left[\frac{1}{b-a} \int_a^b \Phi(t)\, dt - \Phi\left(\frac{a+b}{2}\right) \right].$$

Now we consider another functional depending on the weights

$$K(w; \Phi, f) := \frac{J(w; \Phi, f)}{\int_\Omega w\, d\mu} = \frac{\int_\Omega w(\Phi \circ f)\, d\mu}{\int_\Omega w\, d\mu} - \Phi\left(\frac{\int_\Omega wf\, d\mu}{\int_\Omega w\, d\mu}\right) \geq 0$$

and the composite functional

$$L(w; \Phi, f) := \left(\int_\Omega w\, d\mu \right) \ln\left[K(w; \Phi, f) + 1\right] \geq 0,$$

where $\Phi : I \to \mathbb{R}$ is a continuous convex function on the interval of real numbers I and $f : \Omega \to \mathbb{R}$ is μ-measurable and such that $f, \Phi \circ f \in L_w(\Omega, \mu)$.

Theorem 5. *With the assumptions of Theorem 4, L is a superadditive and monotonic nondecreasing functional of weights.*

Proof. Let $w_i : \Omega \to \mathbb{R}$, with $w_i(x) \geq 0$ for μ – a.e. $x \in \Omega$ and $\int_\Omega w_i\, d\mu > 0, i \in \{1, 2\}$ such that $f, \Phi \circ f \in L_{w_1}(\Omega, \mu) \cap L_{w_2}(\Omega, \mu)$.

Utilising the superadditivity property of J we have

$$L(w_1 + w_2; \Phi, f) \tag{2.11}$$

$$= \left(\int_\Omega (w_1 + w_2)\, d\mu \right) \ln\left[K(w_1 + w_2; \Phi, f) + 1\right]$$

$$= \left(\int_\Omega (w_1 + w_2)\, d\mu \right) \ln\left[\frac{J(w_1 + w_2; \Phi, f)}{\int_\Omega (w_1 + w_2)\, d\mu} + 1 \right]$$

$$\geq \left(\int_\Omega (w_1 + w_2)\, d\mu \right) \ln\left[\frac{J(w_1; \Phi, f) + J(w_2; \Phi, f)}{\int_\Omega (w_1 + w_2)\, d\mu} + 1 \right]$$

$$= \left(\int_\Omega (w_1 + w_2)\, d\mu \right)$$

$$\times \ln\left[\frac{\int_\Omega w_1 d\mu \cdot \frac{J(w_1; \Phi, f)}{\int_\Omega w_1 d\mu} + \int_\Omega w_2 d\mu \cdot \frac{J(w_2; \Phi, f)}{\int_\Omega w_2 d\mu}}{\int_\Omega (w_1 + w_2)\, d\mu} + 1 \right]$$

$$= \left(\int_\Omega (w_1 + w_2)\, d\mu \right)$$

$$\times \ln\left[\frac{\int_\Omega w_1 d\mu \cdot \left(\frac{J(w_1; \Phi, f)}{\int_\Omega w_1 d\mu} + 1 \right) + \int_\Omega w_2 d\mu \cdot \left(\frac{J(w_2; \Phi, f)}{\int_\Omega w_2 d\mu} + 1 \right)}{\int_\Omega (w_1 + w_2)\, d\mu} \right]$$

$$:= A.$$

By the weighted arithmetic mean - geometric mean inequality we have

$$\frac{\int_\Omega w_1 d\mu \cdot \left(\frac{J(w_1;\Phi,f)}{\int_\Omega w_1 d\mu} + 1\right) + \int_\Omega w_2 d\mu \cdot \left(\frac{J(w_2;\Phi,f)}{\int_\Omega w_2 d\mu} + 1\right)}{\int_\Omega (w_1 + w_2)\, d\mu}$$

$$\geq \left(\frac{J(w_1;\Phi,f)}{\int_\Omega w_1 d\mu} + 1\right)^{\frac{\int_\Omega w_1 d\mu}{\int_\Omega (w_1+w_2) d\mu}} \left(\frac{J(w_2;\Phi,f)}{\int_\Omega w_2 d\mu} + 1\right)^{\frac{\int_\Omega w_2 d\mu}{\int_\Omega (w_1+w_2) d\mu}},$$

therefore, by taking the logarithm and utilizing the definition of the functional K, we get the inequality

$$A \geq \left(\int_\Omega w_1 d\mu\right) \ln\left(K(w_1;\Phi,f) + 1\right) + \left(\int_\Omega w_2 d\mu\right) \ln\left(K(w_2;\Phi,f) + 1\right) \tag{2.12}$$

$$= L(w_1;\Phi,f) + L(w_2;\Phi,f).$$

Utilising (2.11) and (2.12) we deduce the superadditivity of the functional L as a function of weights.

Since $L(w;\Phi,f) \geq 0$ for any weight w and it is superadditive, by employing a similar argument to the one in the proof of Theorem 4 we conclude that it is also monotonic nondecreasing as a function of weights. $\qquad\square$

The following result provides another refinement and reverse of the Jensen inequality:

Corollary 2. *Let $w_i : \Omega \to \mathbb{R}$ with $w_i(x) \geq 0$ for $\mu - a.e.$ $x \in \Omega$, $\int_\Omega w_i d\mu > 0, i \in \{1,2\}$ and there exists the nonnegative constants γ, Γ such that*

$$0 \leq \gamma \leq \frac{w_2}{w_1} \leq \Gamma < \infty \ \mu\text{-a.e. on } \Omega.$$

If $\Phi : I \to \mathbb{R}$ is a continuous convex function on the interval of real numbers I, $f : \Omega \to \mathbb{R}$ is μ-measurable and such that $f, \Phi \circ f \in L_{w_1}(\Omega,\mu) \cap L_{w_2}(\Omega,\mu)$, then

$$0 \leq \left[\frac{\int_\Omega w_1(\Phi \circ f)\, d\mu}{\int_\Omega w_1 d\mu} - \Phi\left(\frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu}\right) + 1\right]^{\gamma \cdot \frac{(\int_\Omega w_1 d\mu)}{(\int_\Omega w_2 d\mu)}} - 1 \tag{2.13}$$

$$\leq \frac{\int_\Omega w_2(\Phi \circ f)\, d\mu}{\int_\Omega w_2 d\mu} - \Phi\left(\frac{\int_\Omega w_2 f d\mu}{\int_\Omega w_2 d\mu}\right)$$

$$\leq \left[\frac{\int_\Omega w_1(\Phi \circ f)\, d\mu}{\int_\Omega w_1 d\mu} - \Phi\left(\frac{\int_\Omega w_1 f d\mu}{\int_\Omega w_1 d\mu}\right) + 1\right]^{\Gamma \cdot \frac{(\int_\Omega w_1 d\mu)}{(\int_\Omega w_2 d\mu)}} - 1.$$

Proof. Since L is monotonic nondecreasing and positive homogeneous as a function of weights, we have

$$\gamma L(w_1;\Phi,f) \leq L(w_2;\Phi,f) \leq \Gamma L(w_1;\Phi,f)$$

which is equivalent with

$$[K(w_1;\Phi,f)+1]^{\gamma\left(\int_\Omega w_1 d\mu\right)} \leq [K(w_2;\Phi,f)+1]^{\left(\int_\Omega w_2 d\mu\right)}$$
$$\leq [K(w_1;\Phi,f)+1]^{\Gamma\left(\int_\Omega w_1 d\mu\right)}$$

which provides that

$$[K(w_1;\Phi,f)+1]^{\gamma\frac{\left(\int_\Omega w_1 d\mu\right)}{\left(\int_\Omega w_2 d\mu\right)}} - 1 \leq K(w_2;\Phi,f)$$
$$\leq [K(w_1;\Phi,f)+1]^{\Gamma\frac{\left(\int_\Omega w_1 d\mu\right)}{\left(\int_\Omega w_2 d\mu\right)}} - 1.$$

$\square$

Remark 3. Assume that $\mu(\Omega) < \infty$ and let $w : \Omega \to \mathbb{R}$, with $w(x) \geq 0$ for μ – a.e. $x \in \Omega$, $\int_\Omega w d\mu > 0$ and w is essentially bounded, i.e. ess $\inf_{x\in\Omega} w(x)$ and ess $\sup_{x\in\Omega} w(x)$ are finite.

If $\Phi : I \to \mathbb{R}$ is a continuous convex function on the interval of real numbers I, $f : \Omega \to \mathbb{R}$ is μ-measurable and such that $f, \Phi \circ f \in L_w(\Omega, \mu) \cap L(\Omega, \mu)$, then

$$0 \leq \left[\frac{\int_\Omega (\Phi \circ f) d\mu}{\mu(\Omega)} - \Phi\left(\frac{\int_\Omega f d\mu}{\mu(\Omega)}\right) + 1\right]^{\frac{ess\inf_{x\in\Omega} w(x)}{\frac{1}{\mu(\Omega)}\left(\int_\Omega w d\mu\right)}} - 1 \qquad (2.14)$$
$$\leq \frac{\int_\Omega w(\Phi \circ f) d\mu}{\int_\Omega w d\mu} - \Phi\left(\frac{\int_\Omega w f d\mu}{\int_\Omega w d\mu}\right)$$
$$\leq \left[\frac{\int_\Omega (\Phi \circ f) d\mu}{\mu(\Omega)} - \Phi\left(\frac{\int_\Omega f d\mu}{\mu(\Omega)}\right) + 1\right]^{\frac{ess\sup_{x\in\Omega} w(x)}{\frac{1}{\mu(\Omega)}\left(\int_\Omega w d\mu\right)}} - 1.$$

In particular, if $w : [a,b] \to [0,\infty)$ is Lebesgue integrable, then we have the following result related to the Hermite-Hadamard inequality for the convex function $\Phi : [a,b] \to \mathbb{R}$

$$0 \leq \left[\frac{1}{b-a}\int_a^b \Phi(t) dt - \Phi\left(\frac{a+b}{2}\right) + 1\right]^{\frac{ess\inf_{x\in[a,b]} w(x)}{\frac{1}{b-a}\int_a^b w(t)dt}} - 1 \qquad (2.15)$$
$$\leq \frac{\int_a^b w(t)\Phi(t) dt}{\int_a^b w(t) dt} - \Phi\left(\frac{\int_\Omega w(t)t dt}{\int_a^b w(t) dt}\right)$$
$$\leq \left[\frac{1}{b-a}\int_a^b \Phi(t) dt - \Phi\left(\frac{a+b}{2}\right) + 1\right]^{\frac{ess\sup_{x\in[a,b]} w(x)}{\frac{1}{b-a}\int_a^b w(t)dt}} - 1.$$

3. Applications for the Hölder Inequality

It is well known that if $f \in L_p(\Omega, \mu)$, $p > 1$, where the Lebesgue space $L_p(\Omega, \mu)$ is defined by

$$L_p(\Omega, \mu) := \{ f : \Omega \to \mathbb{R}, \ f \text{ is } \mu\text{-measurable and } \int_\Omega |f(x)|^p \, d\mu(x) < \infty \}$$

and $g \in L_q(\Omega, \mu)$ with $\frac{1}{p} + \frac{1}{q} = 1$ then $fg \in L(\Omega, \mu) := L_1(\Omega, \mu)$ and the *Hölder inequality* holds true

$$\int_\Omega |fg| \, d\mu \le \left(\int_\Omega |f|^p \, d\mu \right)^{1/p} \left(\int_\Omega |g|^p \, d\mu \right)^{1/q}.$$

Assume that $p > 1$. If $h : \Omega \to \mathbb{R}$ is μ-measurable, $\mu(\Omega) < \infty$, $|h|, |h|^p \in L_w(\Omega, \mu) \cap L(\Omega, \mu)$, then by (2.9) we have the bounds

$$0 \le \frac{\operatorname{ess\,inf}_{x \in \Omega} w(x)}{\frac{1}{\mu(\Omega)} \int_\Omega w \, d\mu} \left[\frac{1}{\mu(\Omega)} \int_\Omega |h|^p \, d\mu - \left(\frac{1}{\mu(\Omega)} \int_\Omega |h| \, d\mu \right)^p \right] \tag{3.1}$$

$$\le \frac{1}{\int_\Omega w \, d\mu} \int_\Omega w |h|^p \, d\mu - \left(\frac{1}{\int_\Omega w \, d\mu} \int_\Omega w |h| \, d\mu \right)^p$$

$$\le \frac{\operatorname{ess\,sup}_{x \in \Omega} w(x)}{\frac{1}{\mu(\Omega)} \int_\Omega w \, d\mu} \left[\frac{1}{\mu(\Omega)} \int_\Omega |h|^p \, d\mu - \left(\frac{1}{\mu(\Omega)} \int_\Omega |h| \, d\mu \right)^p \right].$$

Proposition 1. *If $f \in L_p(\Omega, \mu)$, $g \in L_q(\Omega, \mu)$ with $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, $\mu(\Omega) < \infty$ and there exists the constants $\delta, \Delta > 0$ and such that*

$$\delta \le |g| \le \Delta \ \mu\text{-a.e on } \Omega, \tag{3.2}$$

then we have

$$0 \le \frac{\delta^q}{\frac{1}{\mu(\Omega)} \int_\Omega |g|^q \, d\mu} \left[\frac{1}{\mu(\Omega)} \int_\Omega \frac{|f|^p}{|g|^q} \, d\mu - \left(\frac{1}{\mu(\Omega)} \int_\Omega \frac{|f|}{|g|^{q-1}} \, d\mu \right)^p \right] \tag{3.3}$$

$$\le \frac{\int_\Omega |f|^p \, d\mu}{\int_\Omega |g|^q \, d\mu} - \left(\frac{\int_\Omega |fg| \, d\mu}{\int_\Omega |g|^q \, d\mu} \right)^p$$

$$\le \frac{\Delta^q}{\frac{1}{\mu(\Omega)} \int_\Omega |g|^q \, d\mu} \left[\frac{1}{\mu(\Omega)} \int_\Omega \frac{|f|^p}{|g|^q} \, d\mu - \left(\frac{1}{\mu(\Omega)} \int_\Omega \frac{|f|}{|g|^{q-1}} \, d\mu \right)^p \right].$$

Proof. The inequalities (3.3) follows from (3.1) by choosing

$$h = \frac{|f|}{|g|^{q-1}} \text{ and } w = |g|^q.$$

The details are omitted. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 4. We observe that for $p = q = 2$ we have from (3.3) the following reverse of the Cauchy-Bunyakovsky-Schwarz inequality

$$0 \leq \delta^2 \mu\left(\Omega\right) \left[\frac{1}{\mu\left(\Omega\right)} \int_\Omega \left|\frac{f}{g}\right|^2 d\mu - \left(\frac{1}{\mu\left(\Omega\right)} \int_\Omega \left|\frac{f}{g}\right| d\mu\right)^2\right] \int_\Omega |g|^2 d\mu \quad (3.4)$$

$$\leq \int_\Omega |g|^2 d\mu \int_\Omega |f|^2 d\mu - \left(\int_\Omega |fg| d\mu\right)^2$$

$$\leq \Delta^2 \mu\left(\Omega\right) \left[\frac{1}{\mu\left(\Omega\right)} \int_\Omega \left|\frac{f}{g}\right|^2 d\mu - \left(\frac{1}{\mu\left(\Omega\right)} \int_\Omega \left|\frac{f}{g}\right| d\mu\right)^2\right] \int_\Omega |g|^2 d\mu,$$

provided that $f, g \in L_2\left(\Omega, \mu\right)$ and g satisfies the bounds (3.2).

Similar results can be stated by utilizing the inequality (2.13), however the details are not presented here.

4. APPLICATIONS FOR f-DIVERGENCE MEASURES

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 [16], Kullback and Leibler [21], Rényi [27], Havrda and Charvat [14], Kapur [19], Sharma and Mittal [29], Burbea and Rao [4], Rao [26], Lin [22], Csiszár [7], Ali and Silvey [1], Vajda [35], Shioya and Da-te [30] and others (see for example [23] and the references therein).

These measures have been applied in a variety of fields such as: anthropology [26], genetics [23], finance, economics, and political science [28], [32], [33], biology [25], the analysis of contingency tables [13], approximation of probability distributions [6], [20], signal processing [17], [18] and pattern recognition [2], [5]. 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 Ω and the $\sigma-$finite measure μ are given. Consider the set of all probability densities on μ to be $\mathcal{P} := \left\{p | p : \Omega \to \mathbb{R}, \ p\left(x\right) \geq 0, \ \int_\Omega p\left(x\right) d\mu\left(x\right) = 1\right\}$.

Csiszár $f-$divergence is defined as follows [8]

$$I_f\left(p, q\right) := \int_\Omega p\left(x\right) f\left[\frac{q\left(x\right)}{p\left(x\right)}\right] d\mu\left(x\right), \ p, q \in \mathcal{P}, \quad (4.1)$$

where f is convex on $(0, \infty)$. It is assumed that $f\left(u\right)$ is zero and strictly convex at $u = 1$. By appropriately defining this convex function, various divergences are derived.

The *Kullback-Leibler divergence* [21] is well known among the information divergences. It is defined as:

$$D_{KL}(p,q) := \int_\Omega p(x) \ln\left[\frac{p(x)}{q(x)}\right] d\mu(x), \quad p,q \in \mathcal{P}, \tag{4.2}$$

where ln is to base e.

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* [15], χ^2-*divergence D_{χ^2}*, $\alpha-$*divergence D_α, Bhattacharyya distance D_B* [3], *Harmonic distance D_{Ha}, Jeffrey's distance D_J* [16], *triangular discrimination D_Δ* [34], etc... They are defined as follows:

$$D_v(p,q) := \int_\Omega |p(x) - q(x)| d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.3}$$

$$D_H(p,q) := \int_\Omega \left|\sqrt{p(x)} - \sqrt{q(x)}\right| d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.4}$$

$$D_{\chi^2}(p,q) := \int_\Omega p(x)\left[\left(\frac{q(x)}{p(x)}\right)^2 - 1\right] d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.5}$$

$$D_\alpha(p,q) := \frac{4}{1-\alpha^2}\left[1 - \int_\Omega [p(x)]^{\frac{1-\alpha}{2}}[q(x)]^{\frac{1+\alpha}{2}} d\mu(x)\right], \quad p,q \in \mathcal{P}; \tag{4.6}$$

$$D_B(p,q) := \int_\Omega \sqrt{p(x)q(x)} d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.7}$$

$$D_{Ha}(p,q) := \int_\Omega \frac{2p(x)q(x)}{p(x)+q(x)} d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.8}$$

$$D_J(p,q) := \int_\Omega [p(x) - q(x)] \ln\left[\frac{p(x)}{q(x)}\right] d\mu(x), \quad p,q \in \mathcal{P}; \tag{4.9}$$

$$D_\Delta(p,q) := \int_\Omega \frac{[p(x) - q(x)]^2}{p(x)+q(x)} d\mu(x), \quad p,q \in \mathcal{P}. \tag{4.10}$$

For other divergence measures, see the paper [19] by Kapur or the book on line [31] by Taneja.

Most of the above distances $(4.2) - (4.10)$, are particular instances of Csiszár $f-$divergence. There are also many others which are not in this class (see for example [31]). For the basic properties of Csiszár $f-$divergence see [8], [9] and [35].

The following result holds:

Proposition 2. *Let $f : (0,\infty) \to \mathbb{R}$ be a convex function with the property that $f(1) = 0$. Assume that $p,q \in \mathcal{P}$ and there exists constants $0 < s < 1 < S < \infty$ such that*

$$s \le \frac{p(x)}{q(x)} \le S \text{ for } \mu\text{-a.e. } x \in \Omega. \tag{4.11}$$

Then we have the inequalities

$$s\left[I_{f\left(\frac{1}{\cdot}\right)}(q,p) - f\left(D_{\chi^2}(p,q)+1\right)\right] \tag{4.12}$$
$$\leq I_f(p,q)$$
$$\leq S\left[I_{f\left(\frac{1}{\cdot}\right)}(q,p) - f\left(D_{\chi^2}(p,q)+1\right)\right].$$

Proof. If we use the inequality (2.6) we get

$$s\left[\int_\Omega qf\left(\frac{q}{p}\right)d\mu - f\left(\int_\Omega \frac{q^2}{p}d\mu\right)\right] \tag{4.13}$$
$$\leq \int_\Omega pf\left(\frac{q}{p}\right)d\mu$$
$$\leq S\left[\int_\Omega qf\left(\frac{q}{p}\right)d\mu - f\left(\int_\Omega \frac{q^2}{p}d\mu\right)\right].$$

Since

$$\int_\Omega \frac{q^2}{p}d\mu = D_{\chi^2}(p,q)+1$$

and

$$\int_\Omega qf\left(\frac{q}{p}\right)d\mu = I_{f\left(\frac{1}{\cdot}\right)}(q,p),$$

then from (4.13) we deduce the desired result (4.12).   □

We notice that $f\left(\frac{1}{\cdot}\right)$ is not always a convex function. However the definition (4.1) can be extended to any measurable function for which the integral is finite. In applications, the examples for which $f\left(\frac{1}{\cdot}\right)$ is a convex function are of interest. Such an example is provided below.

Consider the Kullback-Leibler divergence defined in (4.2). If $p,q \in \mathcal{P}$ such that there exists constants $0 < s < 1 < S < \infty$ with

$$s \leq \frac{p(x)}{q(x)} \leq S \text{ for } \mu\text{-a.e. } x \in \Omega.. \tag{4.14}$$

then we get from (4.12) that

$$s\left[\ln\left(D_{\chi^2}(p,q)+1\right) - D_{KL}(q,p)\right] \tag{4.15}$$
$$\leq D_{KL}(p,q)$$
$$\leq S\left[\ln\left(D_{\chi^2}(p,q)+1\right) - D_{KL}(q,p)\right].$$

Similar results for f-divergence measures can be stated by utilizing the inequality (2.13), however the details are not presented here.

Acknowledgement

The author would like to thank the anonymous referee for valuable suggestion that have been implemented in the final version of the paper.

References

[1] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. R. Stat. Soc., Ser. B*, vol. 28, pp. 131–142, 1966.

[2] M. Ben-Bassat, "f-entropies, probability of error, and feature selection," *Inf. Control*, vol. 39, pp. 227–242, 1978.

[3] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta Math. Soc.*, vol. 35, pp. 99–109, 1943.

[4] J. Burbea and C. Rao, "On the convexity of some divergence measures based on entropy functions," *IEEE Trans. Inf. Theory*, vol. 28, pp. 489–495, 1982.

[5] C. H. Chen, *Statistical pattern recognition*.   Rocelle Park, New York: Hoyderc Book Co., 1973.

[6] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Trans. Inf. Theory*, vol. 14, pp. 462–467, 1968.

[7] I. Csiszar, "Information-type measures of difference of probability distributions and indirect observations," *Stud. Sci. Math. Hung.*, vol. 2, pp. 299–318, 1967.

[8] I. Csiszar, "On topological properties of f-divergences," *Stud. Sci. Math. Hung.*, vol. 2, pp. 329–339, 1967.

[9] I. Csiszár and J. Körner, *Information theory. Coding theorems for discrete memoryless systems*, ser. Probability and Mathematical Statistics.   New York-San Francisco-London: Academic Press, 1981.

[10] S. S. Dragomir, "A Grüss type inequality for isotonic linear functionals and applications," *Demonstr. Math.*, vol. 36, no. 3, pp. 551–562, 2003.

[11] S. S. Dragomir, "Some reverses of the Jensen inequality with applications," *RGMIA Res. Rep. Coll.*, vol. 14, p. Art. 72., 2011.

[12] S. S. Dragomir and N. M. Ionescu, "Some converse of Jensen's inequality and applications," *Rev. Anal. Numér. Théor. Approx.*, vol. 23, no. 1, pp. 71–78, 1994.

[13] D. Gokhale and S. Kullback, *The information in contingency tables*, ser. Statistics: Textbooks and Monographs.   New York, Basel: Marcel Dekker, Inc., 1978, vol. 23.

[14] J. Havrda and F. Charvat, "Quantification method of classification processes. Concept of structural α-entropy," *Kybernetika, Praha*, vol. 3, pp. 30–35, 1967.

[15] E. Hellinger, "Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen." *J. für Math.*, vol. 136, pp. 210–271, 1909.

[16] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. R. Soc. Lond., Ser. A*, vol. 186, pp. 453–461, 1946.

[17] T. T. Kadota and L. A. Shepp, "On the best finite set of linear observables for discriminating two Gaussian signals," *IEEE Trans. Inf. Theory*, vol. 13, pp. 278–284, 1967.

[18] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. Comm. Technology.*, vol. 15, pp. 52–60, 1967.

[19] J. N. Kapur, "A comparative assessment of various measures of directed divergence," *Advances in Management Studies*, vol. 3, pp. 1–16, 1984.

[20] D. Kazakos and T. Cotsidas, "A decision theory approach to the approximation of discrete probability densities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2, pp. 61–67, 1980.

[21] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, pp. 79–86, 1951.

[22] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, 1991.

[23] M. Mei, "The theory of genetic distance and evaluation of human races," *Japan J. Human Genetics*, vol. 23, pp. 341–369, 1978.

[24] C. P. Niculescu, "An extension of Chebyshev's inequality and its connection with Jensen's inequality," *J. Inequal. Appl.*, vol. 6, no. 4, pp. 451–462, 2001.

[25] E. C. Pielou, *Ecological diversity*.   New York: Wiley, 1975.

[26] C. R. Rao, "Diversity and dissimilarity coefficients: A unified approach," *Theor. Popul. Biol.*, vol. 21, pp. 24–43, 1982.

[27] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Stat. Probab.*, vol. 1, 1961, pp. 547–561.

[28] A. Sen, *On economic inequality*.   London: Oxford University Press, 1973.

[29] B. D. Sharma and D. P. Mittal, "New non-additive measures of relative information," *J. Comb. Inf. Syst. Sci.*, vol. 2, pp. 122–132, 1977.

[30] H. Shioya and T. Da-Te, "A generalisation of Lin divergence and the derivative of a new information divergence," *Elec. and Comm. in Japan*, vol. 78, no. 7, pp. 37–40, 1995.

[31] I. J. Taneja, *Generalized Information Measures and Their Applications*.   electronic, 2001. [Online]. Available: www.mtm.ufsc.br/$\sim$taneja/book/book.html

[32] H. Theil, *Economics and information theory*.   Amsterdam: North-Holland, 1967.

[33] H. Theil, *Statistical decomposition analysis. With applications in the social and administrative sciences*, ser. Studies in Mathematical and Managerial Economics.   Amsterdam-Lond: North-Holland Publishing Comp., 1972, vol. 14.

[34] F. Topsoe, "Some inequalities for information divergence and related measures of discrimination," *Res. Rep. Coll., RGMIA*, vol. 2, no. 1, pp. 85–98, 1999.

[35] I. Vajda, *Theory of statistical inference and information*, ser. Theory and Decision Library, Series B: Mathematical and Statistical Methods.   Dordrecht: Kluwer Academic Publishers, 1989, vol. 11.

Author's address

S. S. Dragomir

Mathematics, School of Engineering & Science, Victoria University, PO Box 14428, Melbourne City, MC 8001, Australia, School of Computational & Applied Mathematics, University of the Witwatersrand, Private Bag 3, Johannesburg 2050, South Africa.

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

URL: http://rgmia.org/dragomir