Received 15/02/12

SHARP UPPER AND LOWER BOUNDS FOR THE JENSEN-DIVERGENCE WITH APPLICATIONS

EDER KIKIANTY, SEVER S. DRAGOMIR, ISIA T. DINTOE, AND DAVID SHERWELL

Abstract. Jensen divergence is used to measure the difference between two probability distributions. In this paper, we consider some bounds for generalised Jensen divergence for twice differentiable functions with bounded second derivatives.

1. Introduction

One of the most important applications of probability theory is finding an appropriate measure of distance (or difference) between two probability distributions [5]. A number of these divergence measures have been widely studied and applied by a number of mathematicians, Burbea and Rao [4], Harvda and Charvát [9], J. Lin [10] and others. These measures can be applied in a variety of fields, even outside of mathematics. We refer the readers to Dragomir [5] for the applications in other areas. The *Jensen's divergence* introduced in Burbea and Rao [4] has tremendous application in the field of mathematical biology especially in Bioinformatics [2, 8], usually is utilised to compare two samples of healthy population (control) and diseased population (case) in detecting gene expression for a certain disease. One of its particular instances is the widely known divergence measure that is known in the literature as the *Jensen-Shannon divergence*. Shannon introduced it in his seminal paper, "A Mathematical Theory of Communications" [11] and it is used in measuring the similarity between two probability distributions.

Following Burbea and Rao [4], we consider the index of diversity of a multinomial distribution, $x = (x_1, \cdots, x_n)$, $x_i \geq 0$ for all $i \in \{1, \ldots, n\}$, such that $\sum_{i=1}^{n} x_i = 1$ and let us define the *Shannon entropy* as, $H_n(x) = -\sum_{i=1}^{n} x_i \log x_i$ (cf. Shannon [11]). The concavity of $H_n(x)$ provides a decomposition of the total diversity in a mixed distribution $\frac{x+y}{2}$ as

$$H_n\left(\frac{x+y}{2}\right) = \frac{1}{2}\left[H_n(x) + H_n(y)\right] + \mathcal{J}_n(x, y). \tag{1}$$

The first component $2^{-1}\left[H_n(x) + H_n(y)\right]$ in (1) is the average Shannon entropy, or diversity, within the distributions, and the second component

$$\mathcal{J}_n(x, y) = H_n\left(\frac{x+y}{2}\right) - \frac{1}{2}\left[H_n(x) + H_n(y)\right], \tag{2}$$

is the *Jensen difference* on divergence, corresponding to the Shannon entropy $H_n(x)$. $H_n(x)$ is nonnegative (as we consider that there is no negative information, see Aczél and Daróczy [1] for properties of the entropy), vanishes if and only if $x = y$, and thus provides a natural measure of divergence between the distributions x and y. Note that, if $\mathcal{J}_n(x, y)$ is treated as a function of two variables (x, y), then $\mathcal{J}_n(x, y)$ is convex.

2000 *Mathematics Subject Classification.* 26D15, 94A17.

Key words and phrases. divergence measure, Jensen divergence, inequality for real numbers.

By extending (1), we get the decomposition of diversity of k distributions $y_1, \ldots, y_k$ with a vector of *a priori* weights $\pi = (\pi_1, \cdots, \pi_k)$, namely we have

$$H_n \left(\sum_{i=1}^{n} \pi_i y_i \right) = \sum_{i=1}^{n} \pi_i H_n(y_i) + \mathcal{J}_n^{\pi}(y_1, \cdots, y_k), \tag{3}$$

where $y_1, \cdots, y_k$ are n-dimensional vectors belonging to the set

$$S_n = \left\{ (x_1, \cdots, x_n) \ \in I_0^n : \sum_{i=1}^{n} x_i = 1 \right\}, \quad I_0 \equiv (0,1). \quad (4) \tag{4}$$

The *generalised Jensen difference* $\mathcal{J}_n^{\pi}$ in (3), which is nonnegative, is the same as the *mutual information (transformation)* defined in Information Theory as a measure of information on a k-input channel for input distribution $\pi = (\pi_1, \cdots, \pi_k)$. For a discussion of the properties of $\mathcal{J}_n^{\pi}$, see Gallager [7, p. 16]; also Aczél and Daróczy [1, p.196–199]). In biological work, $\mathcal{J}_n^{\pi}$ is defined as to be the information radius on the probability distributions associated with $y_1, \cdots, y_k$, (cf. Sibon [12]).

In this paper we consider a function Φ defined on an interval I of the real line $\mathbb{R}$. Let $n \geq 1$ be a positive integer, following Burbea and Rao [4], we define the Φ-entropy of $x = (x_1, \ldots, x_n) \in I \times \cdots \times I = I^n$ as

$$H_{n,\Phi}(x) = - \sum_{i=1}^{n} \Phi(x_i), \tag{5}$$

and for two vectors $x, y \in I$, define also the Jensen divergence by

$$\mathcal{J}_{n,\Phi}(x,y) = H_{n,\Phi} \left(\frac{x+y}{2} \right) - \frac{1}{2} \left[H_{n,\Phi}(x) + H_{n,\Phi}(y) \right]. \tag{6}$$

Then the *Jensen divergence*, which will be referred to as *$\mathcal{J}$-divergence* between $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in I^n$ (where $n \geq 1$, a positive integer) is

$$\mathcal{J}_{n,\Phi}(x,y) := \sum_{i=1}^{n} \left[\frac{1}{2} (\Phi(x_i) + \Phi(y_i) - \Phi \left(\frac{x_i + y_i}{2} \right) \right].$$

The following theorem concerns the property of the convexity of the *$\mathcal{J}$-divergence* measure which is useful for various applications :

Theorem 1 (Burbea-Rao, 1982, [4])**.** *Let Φ be a C^2 function on the interval of real numbers I. Then $J_{n,\Phi}$ is a convex (concave) on $I^n \times I^n$ if and only if Φ is convex (concave) and $\frac{1}{\Phi''}$ is concave (convex) on I.*

In the recent paper by Dragomir, Dragomir and Sherwell [6], the authors found sharp upper and lower bounds for the Jensen divergence for various classes of functions Φ, including functions of bounded variation, absolutely continuous functions, Lipschitzian continuous functions, convex functions and differential functions. We refer to Section 2, for the details of these results obtained which motivates the new results obtained in this paper.

In this paper, we provide new bounds for the *$\mathcal{J}$-divergence* measure of convex and differentiable functions whose derivatives satisfy some boundedness conditions. This bounds are expressed in terms of known and simpler divergence measures that have been employed in the applications mentioned above.

2. Definitions, notations and previous results

In this section, we provide definitions and notation that will be used in the paper. We also provide some results regarding sharp bounds for generalised Jensen divergence as stated in Dragomir, Dragomir and Sherwell [6].

Throughout the paper, for any real number $r>1$, we define r' to be its Hölder conjugate, that is

$$\frac{1}{r}+\frac{1}{r'}=1.$$

Definition 2 (Bullen [3])**.** If s is an extended real number, the generalized logarithmic mean of order s of two positive numbers x and y is defined by

$$\mathfrak{L}^{[s]}(x,y)=\begin{cases}\left[\frac{1}{s+1}\left(\frac{y^{s+1}-x^{s+1}}{y-x}\right)\right]^{\frac{1}{s}}, & \text{if } s\neq -1,0,\pm\infty;\\ \frac{y-x}{\log y-\log x}, & \text{if } s=-1;\\ \frac{1}{e}\left(\frac{y^y}{x^x}\right)^{\frac{1}{y-x}}, & \text{if } s=0;\\ \max\{x,y\}, & \text{if } s=+\infty;\\ \min\{x,y\}, & \text{if } s=-\infty,\end{cases} \tag{7}$$

and $\mathfrak{L}^{[s]}(x,x)=x$.

This mean is homogeneous and symmetric [3, p. 385]. In particular, there is no loss in generality by assuming $0<x<y$. Note also that

$$\mathfrak{L}^{[s]}(x,y)=\left(\int_x^y[(1-t)x+ty]^s dt\right)^{1/s}$$

for $0<x<y$, and $s\in[1,\infty)$. This mean generalises not only logarithmic mean (when $s=-1$), which is particularly useful in distribution of electrical charge of a conductor, but also, arithmetic mean (when $s=1$) and geometric mean (when $s=-2$).

We recall that a function $f:[a,b]\to\mathbb{R}$ is absolutely continuous on $[a,b]$ if and only if it is differentiable almost everywhere in $[a,b]$, the derivative f' is Lebesgue integrable on this interval and $f(y)-f(x)=\int_x^y f'(t)\,dt$ for any $x,y\in[a,b]$.

We use the following notations for Lebesgue integrable functions:

$$\|g\|_{[x,y],p}:=\left|\int_x^y |g(s)|^p\,ds\right|^{1/p} \text{ if } p\geq 1,\ x,y\in[a,b] \text{ and } g\in L_p[a,b];$$

and for $g\in L_\infty[a,b]$ we denote

$$\|g\|_{[x,y],\infty}:=\begin{cases}\operatorname*{ess\,sup}_{s\in[x,y]}|g(s)|, & \text{if } x<y\\ \operatorname*{ess\,sup}_{s\in[y,x]}|g(s)|, & \text{if } y<x.\end{cases}$$

Theorem 3 (Dragomir, Dragomir and Sherwell [6])**.** *Assume that* $\Phi:[a,b]\to\mathbb{R}$ *is absolutely continuous on* $[a,b]$. *Then we have the bounds*

$$\begin{aligned}&|\mathcal{J}_{n,\Phi}(x,y)|\\ &\leq\frac{1}{2}\times\begin{cases}\sum_{i=1}^n|y_i-x_i|\,\|\Phi'\|_{[x_i,y_i],\infty} & \textit{if } \Phi'\in L_\infty[a,b]\\ \sum_{i=1}^n|y_i-x_i|^{\frac{p-1}{p}}\|\Phi'\|_{[x_i,y_i],p} & \textit{if } \Phi'\in L_p[a,b],p>1\\ \sum_{i=1}^n\|\Phi'\|_{[x_i,y_i],1}\end{cases}\\ &\leq\frac{1}{2}\times\begin{cases}\|\Phi'\|_{[a,b],\infty}\sum_{i=1}^n|y_i-x_i| & \textit{if } \Phi'\in L_\infty[a,b]\\ \|\Phi'\|_{[a,b],p}\sum_{i=1}^n|y_i-x_i|^{\frac{p-1}{p}} & \textit{if } \Phi'\in L_p[a,b],p>1\\ n\|\Phi'\|_{[a,b],1}\end{cases}\end{aligned} \tag{8}$$

for any $x=(x_1,...,x_n)\,,\ y=(y_1,...,y_n)\in[a,b]^n$.

Moreover, if the modulus of the derivative is convex, then we have the inequality

$$|\mathcal{J}_{n,\Phi}(x,y)| \le \frac{1}{4}\sum_{i=1}^{n}|y_i-x_i|\left[\left|\Phi'\left(\frac{x_i+y_i}{2}\right)\right|+\frac{|\Phi'(x_i)|+|\Phi'(y_i)|}{2}\right] \le \frac{1}{4}\sum_{i=1}^{n}|y_i-x_i|\left[|\Phi'(x_i)|+|\Phi'(y_i)|\right] \left(\le \|\Phi'\|_{[a,b],\infty}\,\delta(x,y)\right), \tag{9}$$

for any $x=(x_1,...,x_n)\,,\ y=(y_1,...,y_n)\in[a,b]^n$ *where* $\delta(x,y)=\frac{1}{2}\sum_{i=1}^{n}|y_i-x_i|$

The constant $\frac{1}{4}$ *is best possible in both inequalities.*

For two vectors $x=(x_1,...,x_n)\,,\ y=(y_1,...,y_n)\in I^n$ we say that $x\le y$ if for all $i\in\{1,...,n\}$ we have that $x_i\le y_i$. For $x\le y$, we call the set

$$[x,y]:=\{g=(g_1,...,g_n)\mid \text{with } x_i\le g_i\le y_i \text{ for all } i\in\{1,...,n\}\}$$

the generalized interval generated by x and y.

Theorem 4 (Dragomir, Dragomir and Sherwell [6])**.** *Let* $\Phi: I\to\mathbb{R}$ *be a convex function on the interval* I *of real numbers* $\mathbb{R}$.

(i) *If* $x,\ y,\ z\in I^n$ *are so that* $x\le y\le z$, *then*

$$0\le \mathcal{J}_{n,\Phi}(x,y)+\mathcal{J}_{n,\Phi}(y,z)\le \mathcal{J}_{n,\Phi}(x,z)\,, \tag{10}$$

i.e., $\mathcal{J}_{n,\Phi}$ *is superadditive as a functional of the generalized interval;*

(ii) *If* $x,\ y,\ z,u\in I^n$ *are so that* $x\le y\le z\le u$, *then*

$$0\le \mathcal{J}_{n,\Phi}(y,z)\le \mathcal{J}_{n,\Phi}(x,u)\,, \tag{11}$$

i.e., $\mathcal{J}_{n,\Phi}$ *is monotonic nondecreasing as a functional of the generalized interval.*

When more information about the derivative of the function Φ is available, then we can state the following result as well

Theorem 5 (Dragomir, Dragomir and Sherwell [6])**.** *Let* $\Phi:[a,b]\to\mathbb{R}$ *be a differentiable function on the interval* $[a,b]$ *of real numbers* $\mathbb{R}$.

(i) *If the derivative* Φ' *is of bounded variation on* $[a,b]$, *then*

$$|\mathcal{J}_{n,\Phi}(x,y)| \le \frac{1}{4}\sum_{i=1}^{n}|y_i-x_i|\left|\bigvee_{x_i}^{y_i}(\Phi')\right| \le \frac{1}{4}\bigvee_{a}^{b}(\Phi')\sum_{i=1}^{n}|y_i-x_i| = \frac{1}{2}\bigvee_{a}^{b}(\Phi')\,\delta(x,y) \tag{12}$$

for any $x=(x_1,...,x_n)\,,\ y=(y_1,...,y_n)\in[a,b]^n$. *The constant* $\frac{1}{4}$ *is best possible in both inequalities (12).*

(ii) *If the derivative* Φ' *is* K*-Lipschitzian on* $[a,b]$ *with the constant* $K>0$, *then*

$$|\mathcal{J}_{n,\Phi}(x,y)| \le \frac{1}{8}K\sum_{i=1}^{n}(y_i-x_i)^2=\frac{1}{2}K\mathcal{J}_{n,2}(x,y) \tag{13}$$

for any $x=(x_1,...,x_n)\,,\ y=(y_1,...,y_n)\in[a,b]^n$, *where* $\mathcal{J}_{n,2}(x,y)=\frac{1}{4}\sum_{i=1}^{n}(y_i-x_i)^2$. *The constant* $\frac{1}{8}$ *is best possible in (13).*

3. Main results

In this section we provide some bounds for the generalised Jensen divergence for twice differentiable function Φ, whose second derivative Φ'' is bounded above and below in the following sense:

$$\gamma \leq \frac{t^{2-p}}{p(p-1)}\Phi''(t) \leq \Gamma$$

for some $\gamma < \Gamma$ and $p \in (-\infty, 0) \cup (1, \infty)$; and

$$\delta \leq \frac{t^{2-q}}{q(q-1)}\Phi''(t) \leq \Delta$$

for some $\delta \leq \Delta$, and all $q \in (0, 1)$.

Lemma 6 (Dragomir, Dragomir and Sherwell [6])**.** *Let $\Phi : [a, b] \to \mathbb{R}$ be a differentiable function and the derivative Φ' be absolutely continuous. Then,*

$$|\mathcal{J}_{n,\Phi}(x, y)| \leq \begin{cases} \frac{1}{8}\|\Phi''\|_{[a,b],\infty} \sum_{i=1}^{n} (y_i - x_i)^2, & \text{if } \Phi'' \in L_\infty[a, b]; \\ \frac{\|\Phi''\|_{[a,b],r}}{(r'+1)^{1/r'} 2^{1+1/r'}} \sum_{i=1}^{n} |y_i - x_i|^{1+1/r'}, & \text{if } \Phi'' \in L_{r'}[a, b],\ r > 1. \end{cases} \tag{14}$$

We refer to [6] for the proof of this lemma.

Lemma 7. *Let $\Phi : [a, b] \to \mathbb{R}$ be a twice differentiable function and $0 < a < b < \infty$. If $p \in (-\infty, 0) \cup (1, \infty)$ and there exist constants $\gamma < \Gamma$ such that*

$$\gamma \leq \frac{t^{2-p}}{p(p-1)}\Phi''(t) \leq \Gamma$$

for $t \in [a, b]$. Then,

$$\left\| \left(\Phi - \frac{\gamma + \Gamma}{2}(\cdot)^p \right)'' \right\|_{[a,b],\infty} \leq p(p-1)\frac{\Gamma - \gamma}{2} \max\{a^{p-2}, b^{p-2}\}; \tag{15}$$

and

$$\left\| \left(\Phi - \frac{\gamma + \Gamma}{2}(\cdot)^p \right)'' \right\|_{[a,b],r} \leq p(p-1)\frac{\Gamma - \gamma}{2} \left(\mathfrak{L}^{[(p-2)r]}(a, b) \right)^{p-2}, \quad r > 1, \tag{16}$$

where $\mathfrak{L}^{[s]}$ is the sth generalised logarithmic mean.

Proof. Note that

$$\gamma \leq \frac{t^{2-p}}{p(p-1)}\Phi''(t) \leq \Gamma$$

is equivalent to

$$\gamma p(p-1)t^{p-2} \leq \Phi''(t) \leq \Gamma p(p-1)t^{p-2}$$

since $p(p-1) > 0$. This is also equivalent to

$$\left| \Phi''(t) - p(p-1)\frac{\gamma + \Gamma}{2}t^{p-2} \right| \leq p(p-1)\frac{\Gamma - \gamma}{2}t^{p-2}. \tag{17}$$

We take the supremum of both sides to obtain (15). For $r > 1$, we note that (17) is equivalent to

$$\begin{aligned}\|\Phi''(t) - p(p-1)\frac{\gamma+\Gamma}{2}t^{p-2}\|_{[a,b],r} &= \left(\int_a^b \left|\Phi''(t) - p(p-1)\frac{\gamma+\Gamma}{2}t^{p-2}\right|^r dt\right)^{1/r} \\ &\le p(p-1)\frac{\Gamma-\gamma}{2}\left(\int_a^b t^{r(p-2)}\,dt\right)^{1/r} \\ &= p(p-1)\frac{\Gamma-\gamma}{2}\left(\mathfrak{L}^{[r(p-2)]}(a,b)\right)^{p-2}\end{aligned}$$

which proves (16). □

Theorem 8. *Let $\Phi : [a,b] \to \mathbb{R}$ be a twice differentiable function and $0 < a < b < \infty$. If $p \in (-\infty, 0) \cup (1, \infty)$ and there exist constants $\gamma < \Gamma$ such that*

$$\gamma \le \frac{t^{2-p}}{p(p-1)}\Phi''(t) \le \Gamma$$

for $t \in [a,b]$. Then,

$$\begin{aligned}&\left|\mathcal{J}_{n,\Phi}(x,y) - \frac{\gamma+\Gamma}{2}\mathcal{J}_{n,(\cdot)^p}(x,y)\right| \\ &\le \begin{cases} \frac{1}{16}p(p-1)(\Gamma-\gamma)\max\{a^{p-2}, b^{p-2}\}\sum_{i=1}^n (y_i - x_i)^2, & \text{if } \Phi'' \in L_\infty[a,b]; \\ \frac{p(p-1)(\Gamma-\gamma)}{(r'+1)^{1/r'}2^{2+1/r'}}\mathfrak{L}^{[(p-2)r]}(a,b)\sum_{i=1}^n |y_i - x_i|^{1+1/r'}, & \\ & \text{if } \Phi'' \in L_{r'}[a,b],\ r > 1. \end{cases}\end{aligned}$$

Proof. Since any differentiable function is absolutely continuous, we may employ Lemma 6. Combining this with Lemma 7, we have

$$\begin{aligned}&|\mathcal{J}_{n,\Phi}(x,y)| \\ &\le \begin{cases} \frac{1}{8}\|\Phi''\|_{[a,b],\infty}\sum_{i=1}^n (y_i - x_i)^2, \\ \frac{1}{(r'+1)^{1/r'}2^{1+1/r'}}\|\Phi''\|_{[a,b],r}\sum_{i=1}^n |y_i - x_i|^{1+1/r'}, \end{cases} \\ &\le \begin{cases} \frac{1}{8}p(p-1)\frac{\Gamma-\gamma}{2}\max\{a^{p-2}, b^{p-2}\}\sum_{i=1}^n (y_i - x_i)^2, \\ \frac{1}{(r'+1)^{1/r'}2^{1+1/r'}}p(p-1)\frac{\Gamma-\gamma}{2}\left(\mathfrak{L}^{[(p-2)r]}(a,b)\right)^{p-2}\sum_{i=1}^n |y_i - x_i|^{1+1/r'}, \end{cases}\end{aligned}$$

as desired. □

Lemma 9. *Let $\Phi : [a,b] \to \mathbb{R}$ be a twice differentiable function and $0 < a < b < \infty$. If $q \in (0,1)$ and there exist constants $\delta < \Delta$ such that*

$$\delta \le \frac{t^{2-q}}{q(q-1)}\Phi''(t) \le \Delta$$

for $t \in [a,b]$. Then,

$$\left\|\left(\Phi - \frac{\delta+\Delta}{2}(\cdot)^q\right)''\right\|_{[a,b],\infty} \le q(1-q)\frac{\Delta-\delta}{2}\max\{a^{q-2}, b^{q-2}\}; \tag{18}$$

and

$$\left\| \left(\Phi - \frac{\delta+\Delta}{2} (\cdot)^q \right)'' \right\|_{[a,b],q} \leq q(1-q)\frac{\Delta-\delta}{2} \left(\mathfrak{L}^{[r(q-2)]}(a,b) \right)^{p-2}, \quad r > 1, \tag{19}$$

where $\mathfrak{L}^{[s]}$ *is the sth generalised logarithmic mean.*

Proof. Note that

$$\delta \leq \frac{t^{2-q}}{q(q-1)} \Phi''(t) \leq \Delta$$

is equivalent to

$$\Delta q(q-1)t^{q-2} \leq \Phi''(t) \leq \delta q(q-1)t^{q-2}$$

since $q(q-1) < 0$. This is also equivalent to

$$\left| \Phi''(t) - q(q-1)\frac{\delta+\Delta}{2} t^{q-2} \right| \leq q(q-1)\frac{\delta-\Delta}{2} t^{q-2} = q(1-q)\frac{\Delta-\delta}{2} t^{q-2}.$$

Similarly to the proof of (15) in Lemma 7, we take the supremum of both sides to obtain (18). The proof of (19) follows similarly to the proof of (16). □

Theorem 10. *Let* $\Phi : [a,b] \to \mathbb{R}$ *be a twice differentiable function and* $0 < a < b < \infty$. *If* $q \in (0,1)$ *and there exist constants* $\delta < \Delta$ *such that*

$$\delta \leq \frac{t^{2-q}}{q(q-1)} \Phi''(t) \leq \Delta$$

for $t \in [a,b]$. *Then,*

$$\left| \mathcal{J}_{n,\Phi}(x,y) - \frac{\delta+\Delta}{2} \mathcal{J}_{n,(\cdot)^q}(x,y) \right|$$
$$\leq \begin{cases} \frac{1}{16} q(1-q)(\Delta-\delta)\max\{a^{q-2}, b^{q-2}\} \sum_{i=1}^n (y_i - x_i)^2, & \text{if } \Phi'' \in L_\infty[a,b]; \\ \frac{q(1-q)(\Delta-\delta)}{(r'+1)^{1/r'} 2^{2+1/r'}} \left(\mathfrak{L}^{[(q-2)r]}(a,b) \right)^{p-2} \sum_{i=1}^n |y_i - x_i|^{1+1/r'}, & \\ \qquad\qquad \text{if } \Phi'' \in L_{r'}[a,b],\ r > 1. & \end{cases}$$

We omit the proof of this theorem as it follows in a similar manner to the proof of Theorem 8 (combined with Lemma 9).

Theorem 11. *Let* $\Phi : I \subset (0,\infty) \to \mathbb{R}$ *be a twice differentiable function on* I. *If* $p \in (-\infty,0) \cup (1,\infty)$ *and there exists the constants* $\gamma < \Gamma$ *so that*

$$\gamma \leq \frac{t^{2-p}}{p(p-1)} \Phi''(t) \leq \Gamma \text{ for any } t \in I, \tag{20}$$

then we have

$$\gamma(p-1)\mathcal{J}_{n,p}(x,y) \leq \mathcal{J}_{n,\Phi}(x,y) \leq \Gamma(p-1)\mathcal{J}_{n,p}(x,y) \tag{21}$$

for any $x, y \in I$ *where* $\mathcal{J}_{n,\alpha}$ *is defined as*

$$\mathcal{J}_{n,\alpha}(x,y) := (\alpha-1)^{-1} \sum_{i=1}^n \left[\frac{1}{2}(x_i^\alpha + y_i^\alpha) - \left(\frac{x_i + y_i}{2} \right)^\alpha \right] \text{ for } \alpha \neq 1. \tag{22}$$

If $q \in (0,1)$ *and there exists the constants* $\delta < \Delta$ *so that*

$$\delta \leq \frac{t^{2-q}}{q(q-1)} \Phi''(t) \leq \Delta \text{ for any } t \in I \tag{23}$$

then we also have

$$\delta(q-1)\mathcal{J}_{n,q}(x,y) \geq \mathcal{J}_{n,\Phi}(x,y) \geq \Delta(q-1)\mathcal{J}_{n,q}(x,y) \text{ for any } x, y \in I^n. \tag{24}$$

Proof. We consider the auxiliary function $g_{\gamma,p} : I \to \mathbb{R}$ defined by

$$g_{\gamma,p}(t) := \Phi(t) - \gamma t^p,$$

where $p \in (-\infty, 0) \cup (1, \infty)$. We observe that $g_{\gamma,p}$ is twice differentiable on I and the second derivative is given by

$$g''_{\gamma,p}(t) = p(p-1)t^{p-2}\left[\frac{t^{2-p}}{p(p-1)}\Phi''(t) - \gamma\right], \text{for any } t \in I.$$

Utilising the condition (20) and since $p(p-1)t^{p-2} > 0$ for $t \in I$ we deduce that $g''_{\gamma,p}(t) \geq 0$ for any $t \in I$ which means that $g_{\gamma,p}$ is convex on I. Since for a convex function $g : I \to \mathbb{R}$ we have that $\mathcal{J}_{n,g}(x,y) \geq 0$, then we can write that

$$\begin{aligned}
0 &\leq \mathcal{J}_{n,g_{\gamma,p}}(x,y) \\
&= \sum_{i=1}^{n}\left[\frac{g_{\gamma,p}(x_i) + g_{\gamma,p}(y_i)}{2} - g_{\gamma,p}\left(\frac{x_i+y_i}{2}\right)\right] \\
&= \sum_{i=1}^{n}\left[\frac{\Phi(x_i) + \Phi(y_i)}{2} - \Phi\left(\frac{x_i+y_i}{2}\right)\right] - \gamma\sum_{i=1}^{n}\left[\frac{x_i^p + y_i^p}{2} - \left(\frac{x_i+y_i}{2}\right)^p\right] \\
&= \mathcal{J}_{n,\Phi}(x,y) - \gamma(p-1)\mathcal{J}_{n,p}(x,y),
\end{aligned}$$

and the first inequality in (21) is proved.
To prove the second inequality in (21) we consider the auxiliary function

$$g_{\Gamma,p}(x,y) := \Gamma t^p - \Phi(t), \text{ for } t \in I$$

for which we perform a similar argument. The details are omitted.
Now, if $q \in (0,1)$ and if we consider the auxiliary function

$$\psi_{\delta,q}(x,y) := \Phi(t) - \delta t^q,$$

then ψ is twice differentiable and

$$\psi''_{\delta,q}(x,y) = t^{q-2}q(q-1)\left[\frac{t^{2-q}\Phi''(t)}{q(q-1)} - \delta\right] \leq 0 \text{ for any } t \in I,$$

since $q \in (0,1)$. Therefore $\psi_{\delta,q}$ is concave on I which implies that $\mathcal{J}_{n,\psi_{\delta,q}}(x,y) \leq 0$, for any $x, y \in I^n$, and, as above, we obtain

$$\begin{aligned}
\mathcal{J}_{n,\Phi}(x,y) &\leq \delta\sum_{i=0}^{n}\left[\frac{x_i^q + y_i^q}{2} - \left(\frac{x_i+y_i}{2}\right)^q\right] \\
&= \delta(q-1)\mathcal{J}_{n,q}(x,y).
\end{aligned}$$

The second inequality in (24) follows by considering the auxiliary function

$$\psi_{\Delta,q}(x,y) := \Delta t^q - \Phi''(t),$$

and we omit the details. This completes the proof. □

Theorem 12. *Let $\Phi : I \subset (0,\infty) \to \mathbb{R}$ be a twice differentiable function on I. If there exist the constants $\omega < \Omega$ such that*

$$\omega \leq t^2\Phi''(t) \leq \Omega \quad \textit{for any } t \in I, \tag{25}$$

then we have the bounds

$$\omega\mathcal{J}_{n,0}(x,y) \leq \mathcal{J}_{n,\Phi}(x,y) \leq \Omega\mathcal{J}_{n,0}(x,y) \textit{ for any } x,y \in I^n, \tag{26}$$

where

$$\mathcal{J}_{n,0}(x,y) := \sum_{i=1}^{n}\left[\log\left(\frac{x_i+y_i}{2}\right) - \frac{1}{2}(\log x_i + \log y_i)\right]$$
$$\geq 0.$$

If there exists the constants $\lambda < \Lambda$ such that

$$\lambda \leq t\Phi''(t) \leq \Lambda \quad \textit{for any } t \in I, \tag{27}$$

then we have the bounds

$$\lambda \mathcal{J}_{n,1}(x,y) \leq \mathcal{J}_{n,\Phi}(x,y) \leq \Lambda \mathcal{J}_{n,1}(x,y) \textit{ for any } x,y \in I^n, \tag{28}$$

where $J_{n,1}$ is the Jensen-Shannon divergence

$$\mathcal{J}_{n,1}(x,y) := \sum_{i=1}^{n}\left[\frac{x_i\log x_i + y_i \log y_i}{2} - \frac{x_i+y_i}{2}\log\left(\frac{x_i+y_i}{2}\right)\right]$$
$$\geq 0.$$

Proof. Consider the auxiliary function $g_{\omega,0} : I \to \mathbb{R}$ with

$$g_{\omega,0}(t) := \Phi(t) + \omega \log t.$$

We observe that $g_{\omega,0}$ is twice differentiable and by (25),

$$g''_{\omega,0}(t) = t^{-2}(t^2\Phi''(t) - \omega) \geq 0 \text{ for any } t \in I,$$

then we can conclude that $g_{\omega,0}$ is a convex function on I. Therefore we have $\mathcal{J}_{n,g_{\omega,0}}(x,y) \geq 0$ for any $x,y \in I^n$ which implies that

$$\begin{aligned} 0 &\leq \mathcal{J}_{n,\Phi}(x,y) + \omega\sum_{i=1}^{n}\left[\frac{\log x_i + \log y_i}{2} - \log\left(\frac{x_i+y_i}{2}\right)\right] \\ &= \mathcal{J}_{n,\Phi}(x,y) - \omega \mathcal{J}_{n,0}(x,y) \end{aligned}$$

and the first inequality in (2.4) is proved.
Now consider the auxiliary function $g_{\Omega,0} : I \to \mathbb{R}$, with

$$g_{\Omega,0}(t) := -\Omega \log t - \Phi(t), \; t \in I.$$

Then

$$g''_{\Omega,0}(t) = t^{-2}\left(\Omega - t^2\Phi''(t)\right), \quad t \in I,$$

which, by (25) is also a convex function on I. By a similar argument we then deduce the second part of (26).
To prove the second part of the theorem, consider the auxiliary function $g_{\lambda,1} : I \to \mathbb{R}$,

$$g_{\lambda,1}(t) := \Phi(t) - \lambda t \log t, \quad t \in I.$$

We observe that $g_{\lambda,1}$ is twice differentiable and

$$g''_{\lambda,1}(t) := \Phi''(t) - \frac{1}{t}\lambda, \; t \in I.$$

Since, by (27) we have

$$g''_{\lambda,1}(t) = t^{-1}(t\Phi''(t) - \lambda) \geq 0, \quad \text{for any } t \in I,$$

then we can conclude that $g''_{\lambda,1}$ is a convex function on I.
The proof now follows along the lines outlined above and the first part of (28) is proved.
The second part of (28) also follows by employing the auxiliary function $g_{\Lambda,1} : I \to \mathbb{R}$,

$$g_{\Lambda,1}(t) := \Lambda t \log t - \Phi(t), \; t \in I.$$

The theorem is thus proved. □

4. Applications to some elementary functions

We recall the following definitions:

$$\mathcal{J}_{n,\Phi}(x,y) := \sum_{i=1}^{n} \left[\frac{1}{2}(\Phi(x_i) + \Phi(y_i) - \Phi\left(\frac{x_i + y_i}{2}\right) \right],$$

$$\mathcal{J}_{n,\alpha}(x,y) := (\alpha - 1)^{-1} \sum_{i=1}^{n} \left[\frac{1}{2}(x_i^\alpha + y_i^\alpha) - \left(\frac{x_i + y_i}{2}\right)^\alpha \right] \text{ for } \alpha \neq 1,$$

$$\mathcal{J}_{n,0}(x,y) := \sum_{i=1}^{n} \left[\log\left(\frac{x_i + y_i}{2}\right) - \frac{1}{2}(\log x_i + \log y_i) \right],$$

and

$$\mathcal{J}_{n,1}(x,y) := \sum_{i=1}^{n} \left[\frac{x_i \log x_i + y_i \log y_i}{2} - (x_i + y_i)\log\left(\frac{x_i + y_i}{2}\right) \right].$$

We consider the approximation of Jensen divergence as discussed in the previous sections for some elementary functions.

(1) We consider the function $\Phi(t) = e^{-t}$ for $t \in [a,b] \subset [0,1]$ and have the bounds

$$\begin{aligned} a^2 e^{-a} \mathcal{J}_{n,0}(x,y) \le \mathcal{J}_{n,\Phi}(x,y) \le b^2 e^{-b} \mathcal{J}_{n,0}(x,y); \\ a e^{-a} \mathcal{J}_{n,1}(x,y) \le \mathcal{J}_{n,\Phi}(x,y) \le b e^{-b} \mathcal{J}_{n,1}(x,y); \\ \frac{1}{2} e^{-b} \mathcal{J}_{n,2}(x,y) \le \mathcal{J}_{n,\Phi}(x,y) \le \frac{1}{2} e^{-a} \mathcal{J}_{n,2}(x,y). \end{aligned}$$

In what follows, we apply these bounds to the above function on the interval $[0.1, 1]$, where $x = (0.2, 0.25, 0.3, \ldots, 1)$, and $y = (1, \ldots, 1)$.

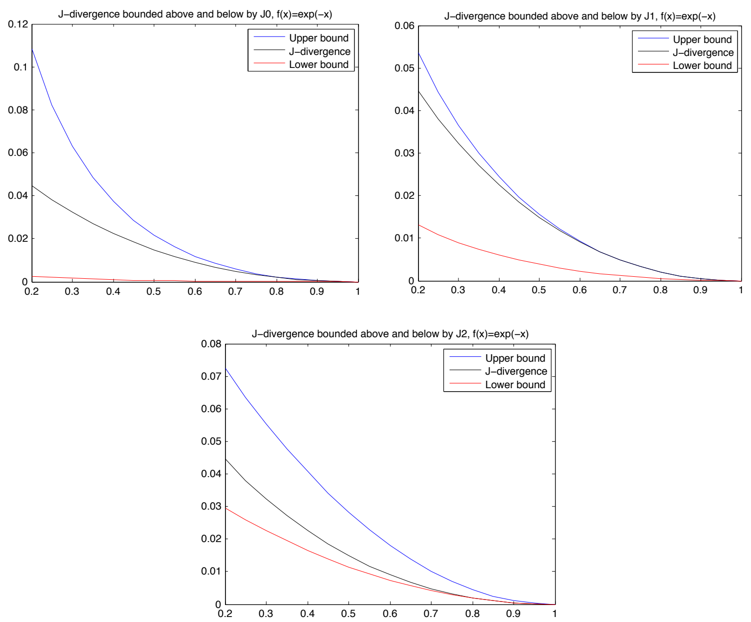

Figure 1. Bounds for generalised Jensen divergence, $\Phi(t) = \exp(-t)$

We also have the following bounds:

$$\left|\mathcal{J}_{n,\Phi}(x,y)-\frac{e^{-a}+e^{-b}}{4}\mathcal{J}_{n,(\cdot)^2}(x,y)\right|$$

$$\leq\begin{cases}\dfrac{1}{16}(e^{-a}-e^{-b})\displaystyle\sum_{i=1}^{n}(y_i-x_i)^2;\\ \dfrac{(e^{-a}-e^{-b})}{(r'+1)^{1/r'}2^{2+1/r'}}e^{-1}\left(\dfrac{b^b}{a^a}\right)^{\frac{1}{b-a}}\displaystyle\sum_{i=1}^{n}|y_i-x_i|^{1+1/r'},\ r>1;\end{cases}$$

(2) We consider the Havrda-Charvát function

$$\Phi_\alpha(t)=\begin{cases}(\alpha-1)^{-1}(t^\alpha-t), & \text{if } \alpha\neq 1,\\ t\log(t), & \text{if } \alpha=1.\end{cases}$$

For $\alpha=1$, we have the following bounds

$$a\mathcal{J}_{n,0}(x,y)\leq\mathcal{J}_{n,\Phi}(x,y)\leq b\mathcal{J}_{n,0}(x,y),\quad \text{for } [a,b]\subset[0,\infty);$$
$$a\mathcal{J}_{n,1}(x,y)\leq\mathcal{J}_{n,\Phi}(x,y)\leq b\mathcal{J}_{n,1}(x,y),\quad \text{for } 0\leq a\leq 1\leq b<\infty.$$

For $\alpha>1$, we have the following bounds

$$\alpha a^\alpha\mathcal{J}_{n,0}(x,y)\leq\mathcal{J}_{n,\Phi}(x,y)\leq\alpha b^\alpha\mathcal{J}_{n,0}(x,y),\quad \text{for } [a,b]\subset[0,\infty);$$
$$\alpha a^{\alpha-1}\mathcal{J}_{n,1}(x,y)\leq\mathcal{J}_{n,\Phi}(x,y)\leq\alpha b^{\alpha-1}\mathcal{J}_{n,1}(x,y),\quad \text{for } [a,b]\subset[0,\infty).$$

In what follows, we apply these bounds to the above function on the interval $[0.1,1]$, where $x=(0.2,0.201,0.202,\ldots,1)$, $y=(1,\ldots,1)$, $\alpha=3/2$.

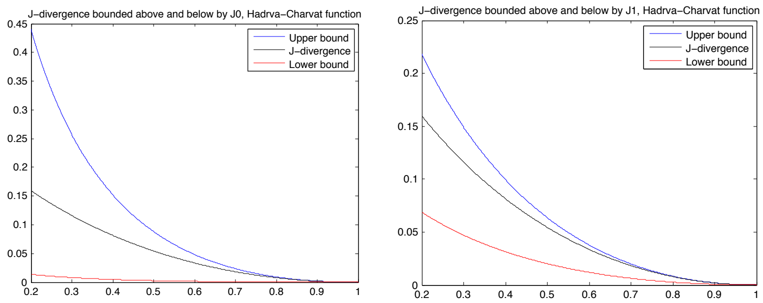

FIGURE 2. BOUNDS FOR GENERALISED JENSEN DIVERGENCE, HAVRDA-CHARVÁT FUNCTION

We also have

$$\left|\mathcal{J}_{n,\Phi}(x,y)-\frac{\gamma+\Gamma}{2}\mathcal{J}_{n,(\cdot)^p}(x,y)\right|$$

$$\leq\begin{cases}\dfrac{1}{16}p(p-1)(\Gamma-\gamma)\max\{a^{p-2},b^{p-2}\}\displaystyle\sum_{i=1}^{n}(y_i-x_i)^2;\\ \dfrac{p(p-1)(\Gamma-\gamma)}{(r'+1)^{1/r'}2^{2+1/r'}}\mathfrak{L}^{[(p-2)r]}(a,b)\displaystyle\sum_{i=1}^{n}|y_i-x_i|^{1+1/r'},\ r>1,\end{cases}$$

and

$$\gamma(p-1)\mathcal{J}_{n,p}(x,y)\leq\mathcal{J}_{n,\Phi}(x,y)\leq\Gamma(p-1)\mathcal{J}_{n,p}(x,y)$$

for $p\in(-\infty,0)\cup(1,\infty)$ where

$$\Gamma=\alpha\frac{b^{\alpha-p}}{p(p-1)},\quad \text{if } [a,b]\subset[0,\infty)$$

and

$$\gamma = \alpha \frac{a^{\alpha-p}}{p(p-1)}, \quad \text{if } [a,b] \subset [0,\infty).$$

Similarly, we have

$$\left| \mathcal{J}_{n,\Phi}(x,y) - \frac{\delta+\Delta}{2} \mathcal{J}_{n,(\cdot)^q}(x,y) \right|$$
$$\leq \begin{cases} \dfrac{1}{16} q(1-q)(\Delta-\delta) \max\{a^{q-2}, b^{q-2}\} \displaystyle\sum_{i=1}^n (y_i - x_i)^2; \\ \dfrac{q(1-q)(\Delta-\delta)}{(r'+1)^{1/r'} 2^{2+1/r'}} \left(\mathfrak{L}^{[(q-2)r]}(a,b) \right)^{p-2} \displaystyle\sum_{i=1}^n |y_i - x_i|^{1+1/r'}, \ r > 1; \end{cases}$$

and

$$\delta(q-1)\mathcal{J}_{n,q}(x,y) \geq \mathcal{J}_{n,\Phi}(x,y) \geq \Gamma(q-1)\mathcal{J}_{n,q}(x,y)$$

for $q \in (0,1)$ where

$$\Delta = \alpha \frac{b^{\alpha-q}}{q(q-1)}, \quad \text{if } [a,b] \subset [0,\infty)$$

and

$$\delta = \alpha \frac{a^{\alpha-q}}{q(q-1)}, \quad \text{if } [a,b] \subset [0,\infty).$$

References

[1] J. Aczél and Z. Daróczy, *On Measures of Information and their Characterization*, New York: Academic (1975).

[2] A.J. Arvey, R.K. Azad, A. Raval and J.G. Lawrence, *Detection of genomic islands via segmental genome heterogeneity*, Nucleic Acids Research **37** (2009), no. 16, 5255–5266.

[3] P.S. Bullen, Handbook of means and their inequalities, Mathematics and its Applications **560**, Kluwer Academic Publishers Group, Dordrecht, 2003.

[4] J. Burbea and C.R. Rao, *On the convexity of some divergence measures based on entropy functions*, IEEE Trans. Inf. Theory. IT-**28**(1982), No. 3, 489-495.

[5] S.S. Dragomir, *Some Reverses of the Jensen Inequality with Applications*, RGMIA Research Report Collection **14**, v14a72, 2011.

[6] S.S. Dragomir, N.M. Dragomir and D. Sherwell, *Sharp Bounds For The Jensen Divergence With Applications*, RGMIA Research Report Collection **14**, v14a47, 2011.

[7] R.G. Gallagher, *Information Theory and Reliable Communication*, New York: Wiley (1968).

[8] R.M. Gmez, O.A. Rosso, R. Berretta, P. Moscato, 2010 *Uncovering Molecular Biomarkers That Correlate Cognitive Decline with the Changes of Hippocampus' Gene Expression Profiles in Alzheimer's Disease*, PLoS ONE **5(4)** (2010): e10153. doi:10.1371/journal.pone.0010153.

[9] M.E. Harvda and F. Charvát, *Quantification Method of Classification Processes: Concept of Structural α-entropy*, Kybernetica **3**, pp. 30-35, 1967.

[10] J. Lin, *Divergence Measures Based On The Shannon Entropy*, IEEE Trans. Inf. Theory. **37**, No. 1, p. 145-151, January (1991).

[11] C.E. Shannon, *A Mathematical Theory Of Communications*, Bell Syst. Tech. J. **27**, pp. 379-423, 623-656, 1948.

[12] R. Sibon, *Information Radius*, Z. Wahrsch'theorie und Verw. Geb. **14**, p.149-160, (1969).

(Eder Kikianty) School of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3 Wits 2050, South Africa.
E-mail address: eder.kikianty@wits.ac.za

(Sever S. Dragomir) School of Engineering and Science, Victoria University, PO Box 14428, Melbourne 8001, Victoria, Australia.
School of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3 Wits 2050, South Africa.
E-mail address: sever.dragomir@vu.edu.au

(Isia T. Dintoe) School of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3 Wits 2050, South Africa.
E-mail address: isia.dintoe@students.wits.ac.za

(David Sherwell) School of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3 Wits 2050, South Africa.
E-mail address: david.sherwell@wits.ac.za