Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/080609

Approximations of the Jensen Divergence

E. Kikianty[1,4,*], *S. S. Dragomir*[2,4], *L. Ndlovu*[3,4] *and D. Sherwell*[4]

[1] Department of Mathematics, University of Johannesburg, PO Box 524, Auckland Park 2006, South Africa
[2] School of Engineering and Science, Victoria University, PO Box 14428, Melbourne 8001, Australia
[3] Standard Bank of South Africa, 5 Simmonds Street, Johannesburg 2001, South Africa
[4] School of Computational and Applied Mathematics, University of the Witwatersrand, Private Bag 3, WITS 2050, South Africa

Received: 10 Oct. 2013, Revised: 8 Jan. 2014, Accepted: 9 Jan. 2014
Published online: 1 Nov. 2014

Abstract: The Jensen divergence is used to measure the difference between two probability distributions. This divergence has been generalised to allow the comparison of more than two distributions. In this paper, we consider some bounds for generalised Jensen divergence for m-time differentiable functions.

Keywords: divergence measure, Jensen divergence, inequality for real numbers.

1. Introduction

In probability theory and statistics, the Jensen divergence is used to measure the difference between two probability distributions. In Burbea and Rao [1], a generalisation of the Jensen divergence is considered to allow the comparison of more than two distributions. If Φ is a function defined on an interval I of the real line $\mathbb{R}$, the *(generalised) Jensen divergence* between two elements $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in I^n (where $n \geq 1$) is given by the following equation, (cf. Burbea and Rao [1])

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

for all $x, y \in I^n \times I^n$. Several measures have been proposed to quantify the difference (also known as the *divergence*) of two (or more) probability distributions. We refer to Grosse et. al. [2], Kullback and Leibler [3], and Csiszar [4] for further references.

These measures can be applied in a variety of fields, for example in fuzzy information systems [5]. The *Jensen divergence* has tremendous applications in the fields of Bioinformatics [7], [8], where it is usually utilised to compare two samples of healthy population (control) and diseased population (case) in detecting gene expression for a certain disease. We refer the readers to Dragomir [6] for applications in other areas.

Recently, Dragomir, Dragomir and Sherwell [10] obtained several sharp bounds for the Jensen divergence, for different classes of functions. We refer the readers to Section 2 for the detail of these results. In the same spirit, we present bounds for m-time differentiable functions in this paper (cf. Sections 3 and 4). Lastly, we apply these bounds for elementary functions in Section 5.

2. Definitions, notation and previous results

In this section, we provide some definitions and notation that are used in the text, and also provide some previous results related to the Jensen divergence. Throughout the paper, we denote p' to be the Hölder conjugate of a real number $1 < p < \infty$, that is, when p' satisfies $1/p + 1/p' = 1$.

We use the following notation for Lebesgue integrable functions. Let $a, b, u, v \in \mathbb{R}$ and without loss of generality, let us assume that $a \leq u \leq v < b$. We denote

$$\|g\|_{[u,v],p} := \left(\int_u^v |g(s)|^p \, ds \right)^{1/p}$$

if $p \geq 1$, $u, v \in [a, b]$ and $g \in L_p[a, b]$. For $g \in L_\infty[a, b]$ we denote

$$\|g\|_{[u,v],\infty} := \operatorname*{ess\,sup}_{s \in [u,v]} |g(s)|.$$

* Corresponding author e-mail: ekikianty@uj.ac.za, eder.kikianty@gmail.com

If $p_1,...,p_k \geq 0$ $(k \geq 2)$ denote the probability distribution satisfying the usual constraints $\sum_{j=1}^{k} p_j = 1$, then the Jensen divergence of the probability distributions is defined by [1]

$$\mathcal{J}^{p}_{n,\Phi}(y^1,...,y^k) := \sum_{i=1}^{n}\left[\sum_{j=1}^{k} p_j \Phi(y_i^j) - \Phi\left(\sum_{j=1}^{k} p_j y_i^j\right)\right]$$

for all $(y^1,...,y^k) \in I^n \times ... \times I^n$ with $y^j = (y_1^j,\dots,y_n^j)$ for $j = 1,\dots,k$. In information theory, $\mathcal{J}_n^p$ defines the measure of information on k-input channel for input distribution $p = p_1,...,p_k$. It also expresses the amount of information supplied by the data for discrimination of these distributions. The divergence $\mathcal{J}_{n,1}$, written as

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

is also known as the *Jensen-Shannon divergence* [9].

Considering the Jensen divergence defined above, we state the following well-known theorem for convex and concave functions.

Theorem 1(Burbea and Rao [1]). *Let Φ be a C^2 function defined on interval I of real numbers. Then $J_{n,\Phi}$ is convex (concave) on $I^n \times I^n$ if and only if Φ is convex (concave) and $1/\Phi''$ is concave (convex) on I. Further, in this case $J^p_{n,\Phi}$ is also convex (concave) on I^{nk} for any given probability distribution p.*

Definition 1. A function $f : [a,b] \to \mathbb{R}$ is absolutely continuous on $[a,b]$ if and only if f is differentiable almost everywhere on $[a,b]$, the derivative f' is Lebesgue integrable on $[a,b]$ and $f(v) - f(u) = \int_u^v f'(t)\,dt$ for any $u,v \in [a,b]$.

Theorem 2(Dragomir, Dragomir, and Sherwell [10]). *Assume that $\Phi : [a,b] \to \mathbb{R}$ is absolutely continuous on $[a,b]$. Then we have the bounds*

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

for any $x = (x_1,...,x_n)$, $y = (y_1,...,y_n) \in [a,b]^n$. The constant $1/4$ 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 $x \leq y$ if for all $i \in \{1,...,n\}$ we have that $x_i \leq y_i$. For $x \leq y$, we call the set,

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

the generalised interval generated by x and y.

Theorem 3(Dragomir, Dragomir, and Sherwell [10]). *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 $x \leq y \leq z$, then

$$0 \leq \mathcal{J}_{n,\Phi}(x,y) + \mathcal{J}_{n,\Phi}(y,z) \leq \mathcal{J}_{n,\Phi}(x,z), \tag{2}$$

i.e., $\mathcal{J}_{n,\Phi}$ is super-additive as a functional of the generalised interval;

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

$$0 \leq \mathcal{J}_{n,\Phi}(y,z) \leq \mathcal{J}_{n,\Phi}(x,u), \tag{3}$$

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

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

Theorem 4(Dragomir, Dragomir, and Sherwell [10]). *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

$$|\mathcal{J}_{n,\Phi}(x,y)| \leq \frac{1}{4}\sum_{i=1}^{n}|y_i - x_i|\left|\bigvee_{x_i}^{y_i}(\Phi')\right| \quad (4)$$

$$\leq \frac{1}{4}\bigvee_{a}^{b}(\Phi')\sum_{i=1}^{n}|y_i - x_i| \quad (5)$$

for any $x = (x_1,...,x_n)$, $y = (y_1,...,y_n) \in [a,b]^n$. The constant $1/4$ is best possible in both inequalities (4) and (5).

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

$$|\mathcal{J}_{n,\Phi}(x,y)| \leq \frac{1}{8}L\sum_{i=1}^{n}(y_i - x_i)^2 \quad (6)$$

$$= \frac{1}{2}L\,\mathcal{J}_{n,2}(x,y)$$

for any $x = (x_1,...,x_n)$, $y = (y_1,...,y_n) \in [a,b]^n$. The constant $1/8$ is best possible in (6).

3. Approximations of Jensen divergence

In this section, we provide some approximations for the following Jensen divergence

$$\mathcal{J}_{1,f}(a,b) = \frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right)$$

for some classes of f which will be used to approximate the generalised Jensen divergence in the later section. We first consider the above Jensen divergence for absolutely continuous functions, and 'weaken' the condition to the case of functions of bounded variation. The results in this section will be used to approximate the Jensen divergence for $\mathcal{J}_{n,\Phi}$ (as defined in Section 1), which we will describe in Section 4.

The following integral identity will be used to obtain an approximation of Jensen divergence. We refer to Cerone, Dragomir and Roumeliotis [11, Lemma 2.1., p. 54].

Lemma 1(Cerone, Dragomir and Roumeliotis [11]). *Let $f : [a,b] \to \mathbb{R}$ be a mapping such that $f^{(m-1)}$ is absolutely continuous on $[a,b]$ we have the identity*

$$\int_a^b f(t)dt = \sum_{k=0}^{m-1}\left[\frac{(b-x)^{k+1}+(-1)^k(x-a)^{k+1}}{(k+1)!}f^{(k)}(x)\right] + \frac{(-1)^m}{m!}\int_a^b K_m(x,t)f^{(m)}(t)dt \quad (7)$$

where the kernel $K_m : [a,b]^2 \to \mathbb{R}$ is given by

$$K_m(x,t) := \begin{cases} (t-a)^m, \text{ if } t \in [a,x] \\ (t-b)^m, \text{ if } t \in (x,b] \end{cases}, \quad x \in [a,b] \quad (8)$$

and m is a natural number, $m \geq 1$.

Corollary 1.*Under the assumptions of Lemma 1, we have the following estimate for the error term in (7)*

$$\left|\frac{(-1)^m}{m!}\int_a^b K_m(x,t)f^{(m)}(t)dt\right| \leq \frac{1}{m!} \times \begin{cases} \dfrac{(x-a)^{m+1}+(b-x)^{m+1}}{m+1}\|f^{(m)}\|_{[a,b],\infty}, \\ \qquad \text{if } f^{(m)} \in L_\infty[a,b]; \\ \dfrac{(x-a)^{m+\frac{1}{p}}+(b-x)^{m+\frac{1}{p}}}{(pm+1)^{\frac{1}{p}}}\|f^{(m)}\|_{[a,b],p'} \\ \qquad \text{if } f^{(m)} \in L_{p'}[a,b], p>1; \\ [(x-a)^m+(b-x)^m]\|f^{(m)}\|_{[a,b],1}, \\ \qquad \text{if } f^{(m)} \in L_1[a,b], \end{cases} \quad (9)$$

for all $x \in [a,b]$.

*Proof.*By Hölder's inequality, we have

$$\left|\frac{(-1)^m}{m!}\int_a^b K_m(x,t)f^{(m)}(t)dt\right| \leq \frac{1}{m!} \times \begin{cases} \int_a^b |K_m(x,t)|dt\,\|f^{(m)}\|_{[a,b],\infty} \\ \left(\int_a^b |K_m(x,t)|^p dt\right)^{1/p}\|f^{(m)}\|_{[a,b],p'}, \quad p>1 \\ \sup_{t\in[a,b]}|K_m(x,t)|\,\|f^{(m)}\|_{[a,b],1}. \end{cases} \quad (10)$$

We evaluate

$$\int_a^b |K_m(x,t)|dt = \int_a^x (t-a)^m dt + \int_x^b (b-t)^m dt$$
$$= \left.\frac{(t-a)^{m+1}}{m+1}\right|_a^x - \left.\frac{(b-t)^{m+1}}{m+1}\right|_x^b$$
$$= \frac{(x-a)^{m+1}+(b-x)^{m+1}}{m+1}$$

which proves the first part of (9). Now,

$$\left(\int_a^b |K_m(x,t)|^p dt\right)^{1/p}$$
$$\leq \left(\int_a^x (t-a)^{pm}dt\right)^{1/p} + \left(\int_x^b (b-t)^{pm}dt\right)^{1/p}$$
$$= \left[\left.\frac{(t-a)^{pm+1}}{pm+1}\right|_a^x\right]^{1/p} + \left[-\left.\frac{(b-t)^{pm+1}}{pm+1}\right|_x^b\right]^{1/p}$$
$$= \frac{(x-a)^{m+1/p}+(b-x)^{m+1/p}}{(pm+1)^{1/p}}$$

which proves the second part of (9). Finally,

$$\sup_{t\in[a,b]} |K_m(x,t)| = \sup_{t\in[a,x]} (t-a)^m + \sup_{t\in(x,b]} (b-t)^m = (x-a)^m + (b-x)^m$$

which completes the proof.

Theorem 5. *Let $f : [a,b] \to \mathbb{R}$ be a mapping such that $f^{(m-1)}$ is absolutely continuous on $[a,b]$. We have the following representation:*

$$\frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right) \tag{11}$$

$$= \sum_{k=1}^{m-1} \frac{(b-a)^k}{2(k+1)!} \left[\left(\frac{1+(-1)^k}{2^k}\right) f^{(k)}\left(\frac{a+b}{2}\right) - f^{(k)}(a) - (-1)^k f^{(k)}(b)\right] \tag{12}$$

$$+\frac{(-1)^m}{2m!(b-a)} \int_a^b C_m(t) f^{(m)}(t)dt,$$

where

$$C_m(t) := \begin{cases} (t-a)^m - (t-b)^m, & \text{if } t \in [a,(a+b)/2]; \\ (t-b)^m - (t-a)^m, & \text{if } t \in ((a+b)/2,b]. \end{cases}$$

Proof. By Lemma 1 we have

$$\int_a^b f(t)dt = (b-a)f(x) + \sum_{k=1}^{m-1} \left[\frac{(b-x)^{k+1} + (-1)^k (x-a)^{k+1}}{(k+1)!} f^{(k)}(x)\right] + \frac{(-1)^m}{m!} \int_a^b K_m(x,t) f^{(m)}(t)dt. \tag{13}$$

Choose $x=a$ in (13) to obtain

$$\int_a^b f(t)dt = (b-a)f(a) + \sum_{k=1}^{m-1} \left[\frac{(b-a)^{k+1}}{(k+1)!} f^{(k)}(a)\right] + \frac{(-1)^m}{m!} \int_a^b (t-b)^m f^{(m)}(t)dt; \tag{14}$$

and choose $x=b$ in (13) to obtain

$$\int_a^b f(t)dt = (b-a)f(b) + \sum_{k=1}^{m-1} \left[\frac{(-1)^k (b-a)^{k+1}}{(k+1)!} f^{(k)}(b)\right] + \frac{(-1)^m}{m!} \int_a^b (t-a)^m f^{(m)}(t)dt. \tag{15}$$

Adding (14) and (15), and divide the sum by 2, we obtain

$$\int_a^b f(t)dt = (b-a)\frac{f(a)+f(b)}{2} + \frac{1}{2}\sum_{k=1}^{m-1} \left[\frac{(b-a)^{k+1}}{(k+1)!}[f^{(k)}(a) + (-1)^k f^{(k)}(b)]\right] + \frac{(-1)^m}{2m!} \int_a^b [(t-a)^m + (t-b)^m] f^{(m)}(t)dt. \tag{16}$$

We also have the following by choosing $x=(a+b)/2$ in (13)

$$\int_a^b f(t)dt = (b-a)f\left(\frac{a+b}{2}\right) + \sum_{k=1}^{m-1} \left(\frac{[1+(-1)^k](b-a)^{k+1}}{2^{k+1}(k+1)!}\right) f^{(k)}\left(\frac{a+b}{2}\right) + \frac{(-1)^m}{m!} \int_a^b M_m(t) f^{(m)}(t)dt, \tag{17}$$

where

$$M_m(t) := \begin{cases} (t-a)^m, & \text{if } t \in [a,(a+b)/2] \\ (t-b)^m, & \text{if } t \in ((a+b)/2,b]. \end{cases}$$

Equating (16) and (17) yields

$$\frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right)$$
$$= \frac{1}{b-a}\left[-\sum_{k=1}^{m-1}\left[\frac{(b-a)^{k+1}}{2(k+1)!}[f^{(k)}(a)+(-1)^k f^{(k)}(b)]\right] + \sum_{k=1}^{m-1} \left(\frac{[1+(-1)^k](b-a)^{k+1}}{2^{k+1}(k+1)!}\right) f^{(k)}\left(\frac{a+b}{2}\right) - \frac{(-1)^m}{2m!}\int_a^b [(t-a)^m + (t-b)^m] f^{(m)}(t)dt + \frac{(-1)^m}{m!}\int_a^b M_m(t) f^{(m)}(t)dt\right]$$
$$= \sum_{k=1}^{m-1} \frac{(b-a)^k}{2(k+1)!} \left[\left(\frac{1+(-1)^k}{2^k}\right) f^{(k)}\left(\frac{a+b}{2}\right) - f^{(k)}(a) - (-1)^k f^{(k)}(b)\right] + \frac{(-1)^m}{2m!(b-a)} \int_a^b C_m(t) f^{(m)}(t)dt,$$

as required.

Corollary 2. *Under the assumptions of Theorem 5, we have the following estimate for the error term in (11):*

$$\left|\frac{(-1)^m}{2m!(b-a)}\int_a^b C_m(x,t)f^{(m)}(t)dt\right|$$
$$\leq \frac{1}{2m!}\begin{cases} \frac{2(b-a)^m}{m+1}\|f^{(m)}\|_{[a,b],\infty}, & \text{if } f^{(m)}\in L_\infty[a,b]; \\ \frac{2(b-a)^{m+1/p-1}}{(pm+1)^{1/p}}\|f^{(m)}\|_{[a,b],p'}, & \text{if } f^{(m)}\in L_{p'}[a,b]; \\ 2(b-a)^{m-1}\|f^{(m)}\|_{[a,b],1}, & \text{if } f^{(m)}\in L_1[a,b]. \end{cases}$$

Proof. Similarly to the proof of Corollary 1, we use Hölder's inequality to estimate the error term (cf. (10)). So, we want to quantify:

$$\begin{aligned}\int_a^b |C_m(t)|dt &= \int_a^b |(t-a)^m-(t-b)^m|dt \\ &\leq \int_a^b (t-a)^m dt + \int_a^b (b-t)^m dt \\ &= \left.\frac{(t-a)^{m+1}}{m+1}\right|_a^b - \left.\frac{(b-t)^{m+1}}{m+1}\right|_a^b \\ &= \frac{2(b-a)^{m+1}}{m+1}.\end{aligned}$$

We also quantify

$$\begin{aligned}&\left(\int_a^b |C_m(t)|^p dt\right)^{1/p} \\ &= \left(\int_a^b |(t-a)^m-(t-b)^m|^p dt\right)^{1/p} \\ &\leq \left(\int_a^b (t-a)^{pm}dt\right)^{1/p} + \left(\int_a^b (b-t)^{pm}dt\right)^{1/p} \\ &= \left[\left.\frac{(t-a)^{pm+1}}{pm+1}\right|_a^b\right]^{1/p} + \left[-\left.\frac{(b-t)^{pm+1}}{pm+1}\right|_a^b\right]^{1/p} \\ &= 2\frac{(b-a)^{m+1/p}}{(pm+1)^{1/p}}.\end{aligned}$$

And finally,

$$\begin{aligned}\sup_{t\in[a,b]}|C_m(t)| &= \sup|(t-a)^m-(t-b)^m| \\ &\leq \sup_{t\in[a,b]}(t-a)^m + \sup_{t\in[a,b]}(b-t)^m \\ &= 2(b-a)^m\end{aligned}$$

which completes the proof.

Remark. For the case of $m=1$, we have the following

$$\begin{aligned}&\frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right) \\ &= -\frac{1}{2(b-a)}\int_a^b C_1(t)f'(t)dt \\ &= \frac{1}{2(b-a)}\left[\int_a^{\frac{a+b}{2}}(a-b)f'(t)dt + \int_{\frac{a+b}{2}}^b (b-a)f'(t)dt\right] \\ &= -\frac{1}{2}\int_a^{\frac{a+b}{2}} f'(t)dt + \frac{1}{2}\int_{\frac{a+b}{2}}^b f'(t)dt\end{aligned}$$

and thus

$$\begin{aligned}&\left|\frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right)\right| \\ &= \left|-\frac{1}{2}\int_a^{\frac{a+b}{2}} f'(t)dt + \frac{1}{2}\int_{\frac{a+b}{2}}^b f'(t)dt\right| \\ &\leq \frac{1}{2}\left|\int_a^{\frac{a+b}{2}} f'(t)dt\right| + \frac{1}{2}\left|\int_{\frac{a+b}{2}}^b f'(t)dt\right| \\ &\leq \frac{1}{2}\int_a^{\frac{a+b}{2}} |f'(t)|dt + \frac{1}{2}\int_{\frac{a+b}{2}}^b |f'(t)|dt \\ &= \frac{1}{2}\int_a^b |f'(t)|dt = \frac{1}{2}\|f'\|_{L^1[a,b]}.\end{aligned}$$

This recaptures the last case in Theorem 2 for $n=1$.

Theorem 6. *Let $f:[a,b]\to\mathbb{R}$ be a function whose m^{th} derivatives $f^{(m)}$ are of locally bounded variation on $[a,b]$. Then,*

$$\begin{aligned}&\frac{f(a)+f(b)}{2} - \left(\frac{a+b}{2}\right) \\ &= \sum_{k=1}^m \left[\frac{(-1)^k+1}{2^{k+1}k!}\right](b-a)^k f^{(k)}\left(\frac{a+b}{2}\right) \\ &+ \frac{1}{2m!}(-1)^{m+1}\int_a^{\frac{a+b}{2}}(s-a)^m\, d(f^{(m)}(s)) \\ &+ \frac{1}{2m!}\int_{\frac{a+b}{2}}^b (b-s)^m\, d(f^{(m)}(s)) \qquad (18)\end{aligned}$$

Proof. We utilise the following Taylor's representation for *m-time* differentiable functions $f:[a,b]\to\mathbb{R}$ whose m^{th} derivatives $f^{(m)}$ are of locally bounded variation on $[a,b]$ (see [8]).

$$f(t) = \sum_{k=0}^m \frac{1}{k!}(t-c)^k f^{(k)}(c) + \frac{1}{m!}\int_c^t (t-s)^m\, d(f^{(m)}(s)) \qquad (19)$$

where *t* and *c* are in $[a,b]$ and the integral in the remainder is taken in the Riemann-Stieltjes sense.

If we choose in (19), $c = (a+b)/2$ and $t = a$, then we get,

$$\begin{aligned} f(a) &= \sum_{k=0}^{m} \frac{1}{k!}\left(\frac{a-b}{2}\right)^k f^{(k)}\left(\frac{a+b}{2}\right) \\ &\quad + \frac{1}{m!}\int_{\frac{a+b}{2}}^{a} (a-s)^m \, d(f^{(m)}(s)) \\ &= \sum_{k=0}^{m} \frac{(-1)^k}{2^k k!}(b-a)^k f^{(k)}\left(\frac{a+b}{2}\right) \\ &\quad + \frac{(-1)^{m+1}}{m!}\int_{a}^{\frac{a+b}{2}} (s-a)^m \, d(f^{(m)}(s)) \\ &= f\left(\frac{a+b}{2}\right) + \sum_{k=1}^{m} \frac{(-1)^k}{2^k k!}(b-a)^k f^{(k)}\left(\frac{a+b}{2}\right) \\ &\quad + \frac{(-1)^{m+1}}{m!}\int_{a}^{\frac{a+b}{2}} (s-a)^m \, d(f^{(m)}(s)). \end{aligned} \tag{20}$$

If we choose in (19), $c = (a+b)/2$ and $t = b$, then we also get,

$$\begin{aligned} f(b) &= \sum_{k=0}^{m} \frac{1}{2^k k!}(b-a)^k f^{(k)}\left(\frac{a+b}{2}\right) \\ &\quad + \frac{1}{m!}\int_{\frac{a+b}{2}}^{b} (b-s)^m \, d(f^{(m)}(s)) \\ &= f\left(\frac{a+b}{2}\right) + \sum_{k=1}^{m} \frac{1}{2^k k!}(b-a)^k f^{(k)}\left(\frac{a+b}{2}\right) \\ &\quad + \frac{1}{m!}\int_{\frac{a+b}{2}}^{b} (b-s)^m \, d(f^{(m)}(s)). \end{aligned} \tag{21}$$

If we add the equality (20) with (21) and divide the sum by 2, then we get,

$$\begin{aligned} \frac{f(a)+f(b)}{2} = \;& f\left(\frac{a+b}{2}\right) \\ & + \sum_{k=1}^{m} \left[\frac{(-1)^k+1}{2^{k+1}k!}\right](b-a)^k f^{(k)}\left(\frac{a+b}{2}\right) \\ & + \frac{1}{2m!}(-1)^{m+1}\int_{a}^{\frac{a+b}{2}} (s-a)^m \, d(f^{(m)}(s)) \\ & + \frac{1}{2m!}\int_{\frac{a+b}{2}}^{b} (b-s)^m \, d(f^{(m)}(s)) \end{aligned}$$

which completes the proof.

Corollary 3. *Under the assumptions of Theorem 6, we have the following estimate for the error term in (18)*

$$\begin{aligned} & \frac{1}{2m!}\left|(-1)^{m+1}\int_{a}^{\frac{a+b}{2}} (s-a)^m \, d(f^{(m)}(s)) \right. \\ & \left. + \int_{\frac{a+b}{2}}^{b} (b-s)^m \, d(f^{(m)}(s))\right| \\ & \le \frac{(b-a)^m}{2^{m+1}m!}\bigvee_a^b (f^{(m)}). \end{aligned} \tag{22}$$

Proof. Note that for any continuous function $p : [\alpha,\beta] \to \mathbb{R}$ and $v : [\alpha,\beta] \to \mathbb{R}$ is of bounded variation, then the Riemann-Stieltjes integral $\int_\alpha^\beta p(t)dv(t)$ exists and

$$\left|\int_\alpha^\beta p(t)dv(t)\right| \le \max_{t\in[\alpha,\beta]} |p(t)| \bigvee_\alpha^\beta (v). \tag{23}$$

Using (23) we have the following

$$\begin{aligned} & \left|\frac{1}{2m!}(-1)^{m+1}\int_{a}^{\frac{a+b}{2}} (s-a)^m \, d(f^{(m)}(s)) \right. \\ & \left. + \frac{1}{2m!}\int_{\frac{a+b}{2}}^{b} (b-s)^m \, d(f^{(m)}(s))\right| \\ & \le \frac{1}{2m!}\left[\max_{t\in[a,(a+b)/2]} (s-a)^m \bigvee_a^{\frac{a+b}{2}} (f^{(m)}) \right. \\ & \left. + \max_{t\in[(a+b)/2,b]} (b-s)^m \bigvee_{\frac{a+b}{2}}^{b} (f^{(m)})\right] \\ & = \frac{1}{2m!}\left[\frac{(b-a)^m}{2^m}\bigvee_a^{\frac{a+b}{2}} (f^{(m)}) + \frac{(b-a)^m}{2^m}\bigvee_{\frac{a+b}{2}}^{b} (f^{(m)})\right] \\ & = \frac{1}{2m!}\left[\frac{(b-a)^m}{2^m}\bigvee_a^b (f^{(m)})\right] = \frac{(b-a)^m}{2^{m+1}m!}\bigvee_a^b (f^{(m)}) \end{aligned}$$

which completes the proof.

Theorem 7(Dragomir [12]). *Let $f : [a,b] \to \mathbb{R}$ be a function whose m^{th} derivatives $f^{(m)}$ are of locally bounded variation on $[a,b]$.*

$$\begin{aligned} f\left(\frac{a+b}{2}\right) &= \frac{f(a)+f(b)}{2} \\ &\quad + \sum_{k=1}^{m} \frac{(b-a)^k}{2^{k+1}k!}\left[f^{(k)}(a) + (-1)^k f^{(k)}(b)\right] \\ &\quad + \int_a^b M_m(t) d\left(f^{(m)}(t)\right), \end{aligned} \tag{24}$$

where

$$M_m(t) = \frac{1}{2m!} \times \begin{cases} \left(\dfrac{a+b}{2} - t\right)^m, & if\, t \in [a, \frac{a+b}{2}] \\ (-1)^m \left(t - \dfrac{a+b}{2}\right)^m, & if\, t \in (\frac{a+b}{2}, b] \end{cases} \tag{25}$$

We refer to [12, Corollary 2] for the proof of this theorem.

Remark. By utilising (23) we have the following bound for the error term in (24)

$$\left|\int_a^b M_m(t) d\left(f^{(m)}(t)\right)\right| \le \frac{(b-a)^m}{2^{m+1}m!}\bigvee_a^b (f^{(m)}). \tag{26}$$

We refer to [12, Corollary 3] for the proof.

4. Approximations of the generalised Jensen divergence

We consider a function $\Phi : I \to \mathbb{R}$ that is *m-time* differentiable $(m \geq 1)$ and the derivative $\Phi^{(m-1)}$ is locally absolutely continuous on I, this means that it is absolutely continuous on any closed subinterval [a,b] of I. For $k = 1, 2, \ldots, m$, we define

$$P_{n,\Phi,k}(x,y) := \sum_{i=1}^{n}(y_i - x_i)^k \Phi^{(k)}\left(\frac{x_i + y_i}{2}\right)$$

$$Q_{n,\Phi,k}(x,y) := \sum_{i=1}^{n}(y_i - x_i)^k \left[\Phi^{(k)}(x_i) + (-1)^k \Phi^{(k)}(y_i)\right]$$

$$E_{n,\Phi,m}(x,y) := \frac{(-1)^m}{2m!} \times \sum_{i=1}^{n}\left[\frac{1}{y_i - x_i}\int_{x_i}^{\frac{x_i+y_i}{2}}\left[(t-x_i)^m - (t-y_i)^m\right]\Phi^{(m)}dt + \frac{1}{y_i - x_i}\int_{\frac{x_i+y_i}{2}}^{y_i}\left[(t-y_i)^m - (t-x_i)^m\right]\Phi^{(m)}dt\right]$$

where $x = (x_1, ..., x_n),\ y = (y_1, ..., y_n) \in I^n$ and the integral above is taken in the sense of a Riemann-Stieltjes. The following representation for the $\mathcal{J}$-divergence can be stated.

Theorem 8.*Let $\Phi : I \to \mathbb{R}$ be a function on I such that the derivative $\Phi^{(m-1)}$ be absolutely continuous on I. Then, Then,*

$$\mathcal{J}_{n,\Phi}(x,y) := \sum_{k=1}^{m-1}\frac{1}{2(k+1)!}\left[\frac{(-1)^k + 1}{2^k}P_{n,\Phi,k}(x,y) - Q_{n,\Phi,k}(x,y)\right] + E_{n,\Phi,k}(x,y)$$

for any vector $x, y \in I^n$.

*Proof.*We employ the result of Theorem 5 for $f \equiv \Phi$, $a = x_i$ and $b = y_i$, $i \in \{1, ..., n\}$ and sum over i from 1 to n, then we deduce the desired representation; and the proof is completed.

Corollary 4.*Under the assumptions of Theorem 8, we have the following estimate:*

$$|E_{n,\Phi,m}(x,y)| \leq \frac{1}{2m!}\begin{cases}\displaystyle\sum_{i=1}^{n}\frac{2|y_i - x_i|^m}{m+1}\max_{i\in\{1,\ldots,n\}}\|\Phi^{(m)}\|_{[x_i,y_i],\infty}, \\ \displaystyle\sum_{i=1}^{n}\frac{2|y_i - x_i|^{m+1/p-1}}{(pm+1)^{1/p}}\max_{i\in\{1,\ldots,n\}}\|\Phi^{(m)}\|_{[x_i,y_i],p'}, \quad p > 1, \\ \displaystyle\sum_{i=1}^{n}2|y_i - x_i|^{m-1}\max_{i\in\{1,\ldots,n\}}\|\Phi^{(m)}\|_{[x_i,y_i],1}.\end{cases}$$

The proof follows by Corollary 2.

We consider now, a function $\Phi : I \to \mathbb{R}$ that is *m-time* differentiable $(m \geq 1)$ and the m^{th} derivative $\Phi^{(m)}$ is of locally bounded variation on I, this means that it is of bounded variation on any closed subinterval [a,b] of I. For k = $1, 2, ..., m$, we recall

$$P_{n,\Phi,k}(x,y) := \sum_{i=1}^{n}(y_i - x_i)^k \Phi^{(k)}\left(\frac{x_i + y_i}{2}\right)$$

and define

$$R_{n,\Phi,m}(x,y) := \frac{1}{2m!}\sum_{i=1}^{n}\left[(-1)^{m+1}\int_{x_i}^{\frac{x_i+y_i}{2}}(t-x_i)^m d(\Phi^{(m)}(t)) + \int_{\frac{x_i+y_i}{2}}^{y_i}(y_i - t)^m d(\Phi^{(m)}(t))\right]$$

where $x = (x_1, ..., x_n),\ y = (y_1, ..., y_n) \in I^n$ and the integral above is taken in the sense of a Riemann-Stieltjes.

The following representation for the $\mathcal{J}$-divergence can be stated.

Theorem 9.*Let $\Phi : I \to \mathbb{R}$ be a m-time differentiable function on I and the m^{th} derivative $\Phi^{(m)}$ be of locally bounded variation on I. Then,*

$$\mathcal{J}_{n,\Phi}(x,y) := \sum_{k=1}^{m}\left[\frac{(-1)^k + 1}{2^{k+1}k!}\right]P_{n,\Phi,k}(x,y) + R_{n,\Phi,m}(x,y) \quad (27)$$

for any vector $x, y \in I^n$.

*Proof.*We employ the result of Theorem 6 for $f \equiv \Phi$, $a = x_i$ and $b = y_i$, $i \in \{1, ..., n\}$ and sum over i from 1 to n, then we deduce the desired representation (27); and the proof is completed.

Corollary 5.*Under the assumptions of Theorem 9, we have the error estimate:*

$$|R_{n,\Phi,m}(x,y)| \leq \frac{1}{2^{m+1}m!}\sum_{i=1}^{n}|y_i - x_i|^m\left|\bigvee_{x_i}^{y_i}(\Phi^{(m)})\right| \leq \frac{1}{2^{m+1}m!}\max_{i\in\{1,\ldots,n\}}\left|\bigvee_{x_i}^{y_i}(\Phi^{(m)})\right|\sum_{i=1}^{n}|y_i - x_i|^m, \quad (28)$$

for any $x, y \in I^m$.

In particular if $x, y \in [a,b]^n \subset I^n$, then we have the simpler bound:

$$|R_{n,\Phi,m}(x,y)| \leq \frac{1}{2^{m+1}m!}\bigvee_{a}^{b}(\Phi^{(m)})\left(\sum_{i=1}^{n}|y_i - x_i|^m\right). \quad (29)$$

The proof follows by Corollary 3.

Corollary 6. *Under the assumptions of Theorem 9 and if the m^{th} derivative $\Phi^{(m)}$ is Lipschitzian with the constant $L_m \geq 0$ on I, then we have the error estimate:*

$$|R_{n,\Phi,m}(x,y)| \leq \frac{L_m}{2^{m+1}(m+1)!} \sum_{i=1}^{n} |y_i - x_i|^{m+1}, \quad (30)$$

for any $x, y \in I^n$.

Proof. It is well known that if $p : [c,d] \to \mathbb{R}$ is a Riemann integrable function and $v : [c,d] \to \mathbb{R}$ is Lipschitzian with constant L, then the Riemann-Stieltjes integral $\int_c^d p(t)\, dv(t)$ exists and,

$$\left| \int_c^d p(t)\, dv(t) \right| \leq L \int_c^d |p(t)|\, dt.$$

Therefore,

$$\begin{aligned}
&|R_{n,\Phi,m}(x,y)| \\
&\leq \frac{1}{2m!} \left\{ \sum_{i=1}^{n} \left[\left| \int_{x_i}^{\frac{x_i+y_i}{2}} (t-x_i)^m\, d(\Phi^{(m)}(t)) \right| \right.\right. \\
&\quad \left.\left. + \left| \int_{\frac{x_i+y_i}{2}}^{y_i} (y_i - t)^m\, d(\Phi^{(m)}(t)) \right| \right] \right\} \\
&\leq \frac{L_m}{2m!} \left\{ \sum_{i=1}^{n} \left[\left| \int_{x_i}^{\frac{x_i+y_i}{2}} (t-x_i)^m\, dt \right| \right.\right. \\
&\quad \left.\left. + \left| \int_{\frac{x_i+y_i}{2}}^{y_i} (y_i - t)^m\, dt \right| \right] \right\} \\
&= \frac{L_m}{2m!} \sum_{i=1}^{n} \left[\frac{|y_i - x_i|^{m+1}}{(m+1)2^{m+1}} + \frac{|y_i - x_i|^{m+1}}{(m+1)2^{m+1}} \right] \\
&= \frac{L_m}{2^{m+1}(m+1)!} \sum_{i=1}^{n} |y_i - x_i|^{m+1}, \quad \text{for any } x, y \in I^n.
\end{aligned}$$

This proves the desired inequality (30).

For $k = 1, 2, \ldots, m$, we now define

$$\begin{aligned}
&Y_{n,\Phi,k}(x,y) \\
&:= -\frac{1}{2m!} \sum_{i=1}^{n} \int_{x_i}^{\frac{x_i+y_i}{2}} \left(\frac{x_i+y_i}{2} - t \right) d(\Phi^{(m)}(t)) \\
&\quad - \frac{1}{2m!} \sum_{i=1}^{n} \int_{\frac{x_i+y_i}{2}}^{y_i} (-1)^m \left(t - \frac{x_i+y_i}{2} \right) d(\Phi^{(m)}(t))
\end{aligned}$$

where $x = (x_1, ..., x_n)$, $y = (y_1, ..., y_n) \in I^n$ and the integral above is taken in the sense of a Riemann-Stieltjes. The following representation for the $\mathcal{J}$-divergence can be stated.

Theorem 10. *Let $\Phi : I \to \mathbb{R}$ be a m-time differentiable function on I and the m^{th} derivative $\Phi^{(m)}$ be of locally bounded variation on I. Then,*

$$\begin{aligned}
\mathcal{J}_{n,\Phi}(x,y) := &-\sum_{i=1}^{n} \sum_{k=1}^{m} \frac{(y_i - x_i)^k}{2^{k+1}k!} \left[\Phi^{(k)}(x_i) \right. \\
&\left. + (-1)^k \Phi^{(k)}(y_i) \right] + Y_{n,\Phi,m}(x,y),
\end{aligned}$$

for any vector $x, y \in I^n$.

The proof follows by Theorem 7.

Corollary 7. *We have the following error estimates*

$$|Y_{n,\Phi,m}(x,y)| \leq \sum_{i=1}^{n} \frac{|y_i - x_i|^m}{2^{m+1}m!} \max_{i\{1,\ldots,n\}} \left[\bigvee_{x_i}^{y_i} (\Phi^{(m)}) \right].$$

The proof follows by Remark 3.

5. Application to some elementary functions

In this section, we consider the approximation of Jensen divergence for some elementary functions.

1. First, we consider the exponential function, i.e. $\Phi(t) = e^t$. We have, from Theorem 8

$$\begin{aligned}
\mathcal{J}_{n,e^t}(x,y) \approx &\sum_{i=1}^{n} \sum_{k=1}^{m} \frac{(y_i - x_i)^k}{2(k+1)!} \\
&\times \left[\frac{1 + (-1)^k}{2^k} e^{\frac{x_i+y_i}{2}} - e^{x_i} - (-1)^k e^{y_i} \right]
\end{aligned}$$

with the remainder $E_{n,e^t,m}(x,y)$ satisfies the bound,

$$\begin{aligned}
&|E_{n,e^t,m}| \\
&\leq \frac{1}{2m!} \times \begin{cases} \displaystyle\sum_{i=1}^{n} \frac{2|y_i - x_i|^m}{m+1} \max_{i \in \{1,\ldots,n\}} e^{y_i}, \\ \displaystyle\sum_{i=1}^{n} \frac{2|y_i - x_i|^{m+1/p-1}}{(pm+1)^{1/p}} \\ \displaystyle\quad \times \max_{i \in \{1,\ldots,n\}} \left(\frac{e^{p' y_i} - e^{p' x_i}}{p'} \right)^{1/p'}, \quad p > 1, \\ \displaystyle\sum_{i=1}^{n} 2|y_i - x_i|^{m-1} \max_{i \in \{1,\ldots,n\}} (e^{y_i} - e^{x_i}). \end{cases}
\end{aligned}$$

Theorem 9 gives us

$$\mathcal{J}_{n,e^t}(x,y) \approx \sum_{i=1}^{n} \sum_{k=1}^{m} \frac{(-1)^k + 1}{2^{k+1}k!} (y_i - x_i)^k e^{\frac{x_i+y_i}{2}}$$

where the remainder, $R_{n,e^t,m}(x,y)$ satisfies the bound,

$$\begin{aligned}
&|R_{n,e^t,m}(x,y)| \\
&\leq \frac{1}{2^{m+1}m!} \max_{i \in \{1,\ldots n\}} (e^{y_i} - e^{x_i}) \sum_{i=1}^{n} |y_i - x_i|^m \\
&\leq \frac{1}{2^{m+1}m!} \max_{i \in \{1,\ldots n\}} (e^{y_i} - e^{x_i}) \left(\sum_{i=1}^{n} |y_i - x_i| \right)^m.
\end{aligned}$$

Finally, Theorem 10 gives us

$$\mathcal{J}_{n,e^t}(x,y) \approx -\sum_{i=1}^{n}\sum_{k=1}^{m} \frac{(y_i-x_i)^k}{2^{k+1}k!}\left[e^{x_i}+(-1)^k e^{y_i}\right]$$

with the remainder $Y_{n,e^t,m}(x,y)$ satisfies the bound,

$$|Y_{n,e^t,m}(x,y)| \le \frac{1}{2^{m+1}m!}\max_{i\in\{1,\ldots,n\}}\left(e^{y_i}-e^{x_i}\right)\sum_{i=1}^{n}|y_i-x_i|^m.$$

2.We now consider the function $\Phi(t) = t^p$, where $p > m$. We have, from Theorem 8

$$\mathcal{J}_{n,t^p}(x,y) \approx \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{(y_i-x_i)^k}{2(k+1)!}\left[\frac{1+(-1)^k}{2^k}\frac{p!}{(p-k)!}\left(\frac{x_i+y_i}{2}\right)^{p-k} - \frac{p!}{(p-k)!}{x_i}^{p-k} - \frac{p!(-1)^k}{(p-k)!}{y_i}^{p-k}\right]$$

with the remainder $E_{n,e^t,m}(x,y)$ satisfies the bound,

$$|E_{n,t^p,m}| \le \frac{p!}{2m!(p-m)!} \times \begin{cases} \displaystyle\sum_{i=1}^{n}\frac{2|y_i-x_i|^m}{m+1}\max_{i\in\{1,\ldots,n\}}\|t^{p-m}\|_{[x_i,y_i],\infty}, \\ \displaystyle\sum_{i=1}^{n}\frac{2|y_i-x_i|^{m+1/p-1}}{(pm+1)^{1/p}}\max_{i\in\{1,\ldots,n\}}\|t^{p-m}\|_{[x_i,y_i],p'}, \quad p>1, \\ \displaystyle\sum_{i=1}^{n}2|y_i-x_i|^{m-1}\max_{i\in\{1,\ldots,n\}}\|t^{p-m}\|_{[x_i,y_i],1}. \end{cases}$$

Theorem 9 gives us

$$\mathcal{J}_{n,t^p}(x,y) \approx \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{(-1)^k+1}{2^{k+1}k!}(x_i-y_i)^k\frac{p!}{(p-k)!}\left(\frac{x_i+y_i}{2}\right)^{p-k}$$

where the remainder, $R_{n,t^p,m}(x,y)$ satisfies the bound,

$$|R_{n,e^t,m}(x,y)| \le \frac{p!}{2^{m+1}m!(p-m-1)!} \times \max_{i\in\{1,\ldots n\}}\|t^{p-m-1}\|_{[x_i,y+i,1]}\sum_{i=1}^{n}|y_i-x_i|^m.$$

Finally, Theorem 10 gives us

$$\mathcal{J}_{n,e^t}(x,y) \approx -\sum_{i=1}^{n}\sum_{k=1}^{m}\frac{(y_i-x_i)^k}{2^{k+1}k!}\frac{p!}{(p-k)!}\left[{x_i}^{p-k}+(-1)^k{y_i}^{p-k}\right]$$

with the remainder $Y_{n,e^t,m}(x,y)$ satisfies the bound,

$$|Y_{n,e^t,m}(x,y)| \le \frac{p!}{2^{m+1}m!(p-m-1)!} \times \max_{i\in\{1,\ldots n\}}\|t^{p-m-1}\|_{[x_i,y+i,1]}\sum_{i=1}^{n}|y_i-x_i|^m.$$

3.We consider the function $\Phi(t) = -\log(t)$, where $t \ge 1$. We have, from Theorem 8

$$\mathcal{J}_{n,t^p}(x,y) \approx \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{(y_i-x_i)^k}{2k(k+1)} \times \left[\frac{(-1)^k+1}{2^k}\left(\frac{x_i+y_i}{2}\right)^{-k} - (-1)^k x_i^{-k} - y_i^{-k}\right]$$

with the remainder $E_{n,e^t,m}(x,y)$ satisfies the bound,

$$|E_{n,t^p,m}| \le \frac{1}{2m} \times \begin{cases} \displaystyle\sum_{i=1}^{n}\frac{2|y_i-x_i|^m}{m+1}\max_{i\in\{1,\ldots,n\}}\|t^{-m}\|_{[x_i,y_i],\infty}, \\ \displaystyle\sum_{i=1}^{n}\frac{2|y_i-x_i|^{m+1/p-1}}{(pm+1)^{1/p}}\max_{i\in\{1,\ldots,n\}}\|t^{-m}\|_{[x_i,y_i],p'}, \quad p>1, \\ \displaystyle\sum_{i=1}^{n}2|y_i-x_i|^{m-1}\max_{i\in\{1,\ldots,n\}}\|t^{-m}\|_{[x_i,y_i],1}. \end{cases}$$

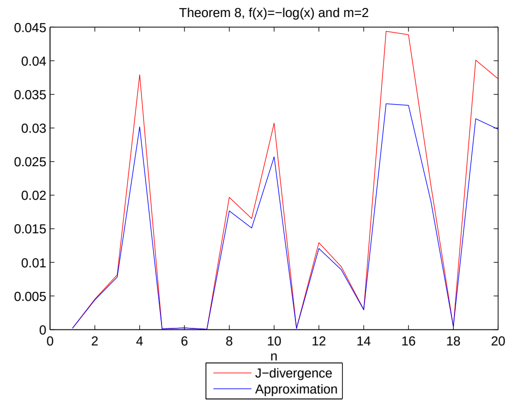
Theorem 9 gives us

$$\mathcal{J}_{n,t^p}(x,y) \approx \sum_{i=1}^{n}\sum_{k=1}^{m}\frac{1+(-1)^k}{2k}\left(\frac{x_i-y_i}{x_i+y_i}\right)^k$$

where the remainder, $R_{n,t^p,m}(x,y)$ satisfies the bound,
$|R_{n,e^t,m}(x,y)| \le \frac{1}{2^{m+1}}\max_{i\in\{1,\ldots n\}}\|t^{-m-1}\|_{[x_i,y+i,1]}\sum_{i=1}^{n}|y_i-x_i|^m.$
Finally, Theorem 10 gives us

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

with the remainder $Y_{n,e^t,m}(x,y)$ satisfies the bound,
$|Y_{n,e^t,m}(x,y)| \le \frac{1}{2^{m+1}}\max_{i\in\{1,\ldots n\}}\|t^{-m-1}\|_{[x_i,y+i,1]}\sum_{i=1}^{n}|y_i-x_i|^m.$
In the following figures, choose $n = 20$, $m = 2,4,6$ for $\Phi(t) = -\log(t)$, $I = [10,20]$. We observe that the approximation in Theorem 9 converges faster than the other two, whilst the approximation in Theorem 8 is the slowest.



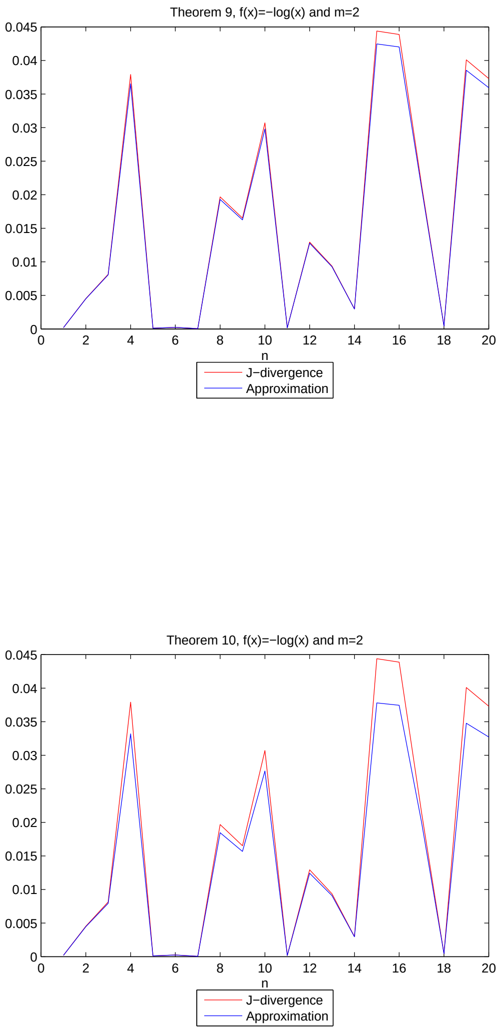


Figure 1: Jensen divergence and its approximation $(m = 2)$

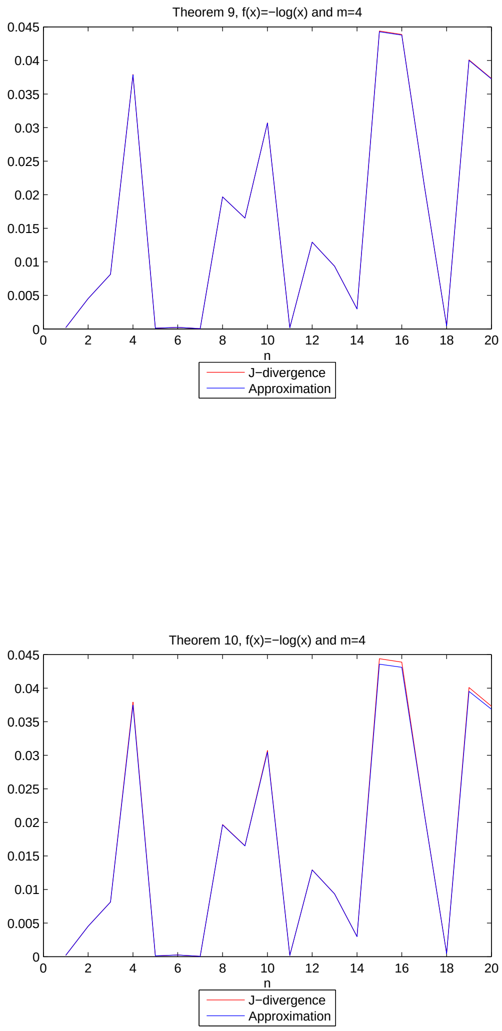
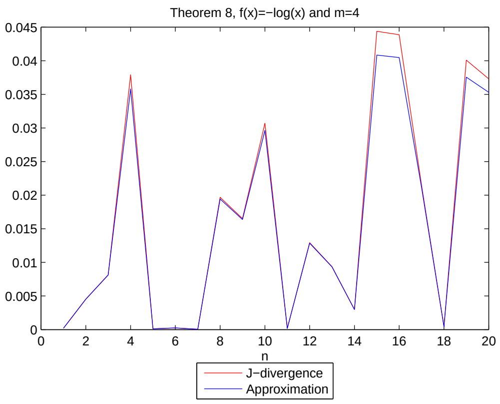
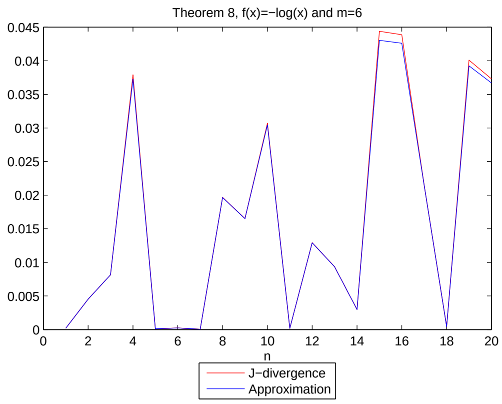
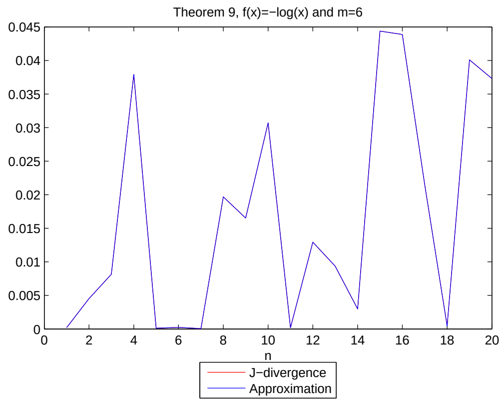


Figure 2: Jensen divergence and its approximation $(m = 4)$







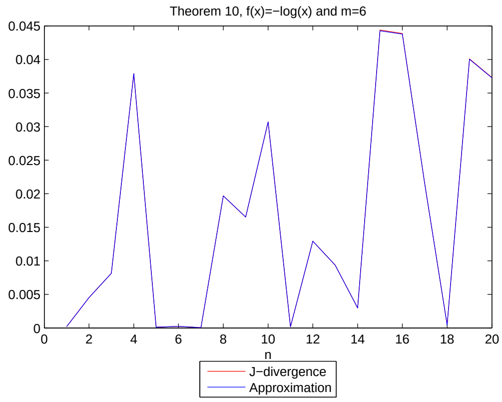


Figure 3: Jensen divergence and its approximation $(m = 6)$

References

[1] J. Burbea and C.R. Rao, On the convexity of some divergence measures based on entropy functions, IEEE Transaction on Information Theory IT, **28**, 489–495 (1982).

[2] I. Grosse, P. Bernaola-Galvan, P. Carpena, R. Roman-Roldan, J. Oliver, and H. E. Stanley, Analysis of symbolic sequences using the Jensen-Shannon divergence, Physical Review E - Statistical, Nonlinear and Soft Matter Physics, **65**, 041905 (2002) http://link.aps.org/doi/10.1103/PhysRevE.65.041905.

[3] S. Kullback and R.A. Leibler, On information and sufficiency, Annals of Mathematical Statistics., **22**, 79–86 (1951).

[4] I. Csiszar, Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar., **2**, 299–318 (1967).

[5] M.L. Menendez, J.A. Pardo and L. Pardo, Some statistical applications of generalized Jensen difference divergence measures for fuzzy information systems, Fuzzy Sets and Systems, **52**, 169–180 (1992).

[6] S.S. Dragomir, Some reverses of the Jensen inequality with applications, RGMIA Research Report Collection, **14**, (2011)

[7] A.J. Arvey, R.K. Azad, A. Raval and J.G. Lawrence, Detection of genomic islands via segmental genome heterogeneity, Nucleic Acids Research, **37**, 5255–5266 (2009).

[8] R.M. Gómez, O.A. Rosso, R. Berretta, P. Moscato, Uncovering molecular biomarkers that correlate cognitive decline with the changes of Hippocampus' gene expression profiles in Alzheimer's disease, PLoS ONE, **5**, (2010) doi:10.1371/journal.pone.0010153.

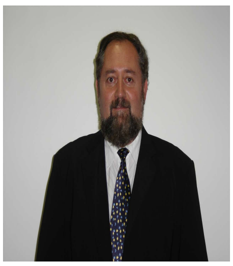
[9] C.E. Shannon, A mathematical theory of communications, Bell System Technical Journal, **27**, 379–423, 623–565 (1948).

[10] S.S. Dragomir, N.M. Dragomir and D. Sherwell, Sharp bounds for the Jensen divergence with applications, RGMIA Research Report Collection, **14**, v14a47 (2011).

[11] P. Cerone, S.S. Dragomir, and J. Roumeliotis, Some Ostrowski type inequalities for n-time differentiable mappings and applications, Demonstratio Mathematica, **32**, 697–712 (1999).

[12] S.S. Dragomir, Approximating real functions which possess n^{th} derivatives of bounded variation and applications, Computers and Mathematics Applications, **56**, 2268–2278 (2008).

Eder Kikianty is a postdoctoral fellow at the Department of Mathematics, University of Johannesburg (South Africa). She received her Bachelor of Science (with Honours) in Mathematics from Institut Teknologi Bandung (Indonesia) in 2005, and her Doctor of Philosophy in Mathematics from Victoria University (Australia) in 2010. Her current research interests include inequality theory, geometry of Banach spaces and Banach lattices.



S. S. Dragomir is a Chair on Mathematical Inequalities at Victoria University, Melbourne Australia and Honorary Professor at the University of Witwatersrand, Johannesburg, South Africa. He is also the Chair of the international Research Group in Mathematical Inequalities & Applications (RGMIA). For details of his publications see his Google Scholar profile.

L. P. Ndlovu is currently a Quantitative Analyst at Standard Bank of South Africa. He received a Bachelor of Science and Master of Science (with Distinction) in Applied Mathematics from the University of the Witwatersrand (South Africa).