CrossMark

About the sharpness of the Jensen inequality

Đilda Pečarić[1], Josip Pečarić[2,3] and Mirna Rodić[2*] (ID)

*Correspondence: mrodic@ttf.hr
[2]Faculty of Textile Technology, University of Zagreb, Zagreb, Croatia
Full list of author information is available at the end of the article

Abstract

The main aim of this paper is to give an improvement of the recent result on the sharpness of the Jensen inequality. The results given here are obtained using different Green functions and considering the case of the real Stieltjes measure, not necessarily positive. Finally, some applications involving various types of f-divergences and Zipf–Mandelbrot law are presented.

MSC: 26D15; 60E05; 94A17

Keywords: Green function; Jensen's inequality; f-divergence; Entropy; Zipf–Mandelbrot law

1 Introduction

The Jensen inequality is one of the most famous and most important inequalities in mathematical analysis.

In [2], some estimates about the sharpness of the Jensen inequality are given. In particular, the difference

$$\int_0^1 \varphi\big(f(x)\big)\,dx - \varphi\bigg(\int_0^1 f(x)\,dx\bigg)$$

is estimated, where φ is a convex function of class C^2.

The authors in [2] expanded $\varphi(f(x))$ around any given value of $f(x)$, say around $c = f(x_0)$, which can be arbitrarily chosen in the domain I of φ, such that $c = f(x_0)$ is in the interior of I, and as their first result, they get the following inequalities:

$$
\begin{aligned}
0 &\le \int_0^1 \varphi\big(f(x)\big)\,dx - \varphi\bigg(\int_0^1 f(x)\,dx\bigg) \\
&\le \frac{1}{2}\big\|\varphi''\big\|_{L^\infty(I_2)} \cdot \|f - c\|_{L^2}^2 + \frac{1}{2}\big\|\varphi''\big\|_{L^\infty(I_2)} \cdot \|f - c\|_{L^1}^2 \\
&= \frac{1}{2}\big\|\varphi''\big\|_{L^\infty(I_2)} \cdot \big[\|f - c\|_{L^2}^2 + \|f - c\|_{L^1}^2\big],
\end{aligned}
$$

where I_2 denotes the domain of φ''.

The main aim of our paper is to give an improvement of that result using various Green functions and considering the case of the real Stieltjes measure, not necessarily positive.

Springer

2 Preliminary results

Consider the Green functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ $(k = 0, 1, 2, 3, 4)$ defined by

$$G_0(t,s) = \begin{cases} \frac{(t-\beta)(s-\alpha)}{\beta-\alpha} & \text{for } \alpha \le s \le t, \\ \frac{(s-\beta)(t-\alpha)}{\beta-\alpha} & \text{for } t \le s \le \beta, \end{cases} \tag{1}$$

$$G_1(t,s) = \begin{cases} \alpha - s & \text{for } \alpha \le s \le t, \\ \alpha - t & \text{for } t \le s \le \beta, \end{cases} \tag{2}$$

$$G_2(t,s) = \begin{cases} t - \beta & \text{for } \alpha \le s \le t, \\ s - \beta & \text{for } t \le s \le \beta, \end{cases} \tag{3}$$

$$G_3(t,s) = \begin{cases} t - \alpha & \text{for } \alpha \le s \le t, \\ s - \alpha & \text{for } t \le s \le \beta, \end{cases} \tag{4}$$

$$G_4(t,s) = \begin{cases} \beta - s & \text{for } \alpha \le s \le t, \\ \beta - t & \text{for } t \le s \le \beta. \end{cases} \tag{5}$$

All these functions are convex and continuous with respect to both s and t.

We have the following lemma (see also [16] and [17]).

Lemma 1 *For every function $\varphi \in C^2[\alpha, \beta]$, we have the following identities*:

$$\varphi(x) = \frac{\beta - x}{\beta - \alpha}\varphi(\alpha) + \frac{x - \alpha}{\beta - \alpha}\varphi(\beta) + \int_\alpha^\beta G_0(x,s)\varphi''(s)\,ds, \tag{6}$$

$$\varphi(x) = \varphi(\alpha) + (x - \alpha)\varphi'(\beta) + \int_\alpha^\beta G_1(x,s)\varphi''(s)\,ds, \tag{7}$$

$$\varphi(x) = \varphi(\beta) + (x - \beta)\varphi'(\alpha) + \int_\alpha^\beta G_2(x,s)\varphi''(s)\,ds, \tag{8}$$

$$\varphi(x) = \varphi(\beta) - (\beta - \alpha)\varphi'(\beta) + (x - \alpha)\varphi'(\alpha) + \int_\alpha^\beta G_3(x,s)\varphi''(s)\,ds, \tag{9}$$

$$\varphi(x) = \varphi(\alpha) + (\beta - \alpha)\varphi'(\alpha) - (\beta - x)\varphi'(\beta) + \int_\alpha^\beta G_4(x,s)\varphi''(s)\,ds, \tag{10}$$

where the functions G_k $(k = 0, 1, 2, 3, 4)$ are defined as before in (1)–(5).

Proof By integrating by parts we get

$$\int_\alpha^\beta G_0(x,s)\varphi''(s)\,ds = \int_\alpha^x G_0(x,s)\varphi''(s)\,ds + \int_x^\beta G_0(x,s)\varphi''(s)\,ds$$

$$= \int_\alpha^x \frac{(x-\beta)(s-\alpha)}{\beta-\alpha}\varphi''(s)\,ds + \int_x^\beta \frac{(s-\beta)(x-\alpha)}{\beta-\alpha}\varphi''(s)\,ds$$

$$= \frac{x-\beta}{\beta-\alpha}\int_\alpha^x (s-\alpha)\varphi''(s)\,ds + \frac{x-\alpha}{\beta-\alpha}\int_x^\beta (s-\beta)\varphi''(s)\,ds$$

$$= \varphi(x) - \frac{\beta-x}{\beta-\alpha}\varphi(\alpha) - \frac{x-\alpha}{\beta-\alpha}\varphi(\beta).$$

The other identities are proved analogously.   □

Remark 1 The result (7) given in the previous lemma represents a special case of the representation of the function φ using the so-called "two-point right focal" interpolating polynomial in the case where $n = 2$ and $p = 0$ (see [1]).

Using the results from the previous lemma, the authors in [16] and [17] gave the uniform treatment of the Jensen-type inequalities, allowing the measure also to be negative. In this paper, we give some further interesting results.

3 Main results

To simplify the notation, for functions g and λ, we denote

$$\overline{g} = \frac{\int_a^b g(x)\, d\lambda(x)}{\int_a^b d\lambda(x)}.$$

Theorem 1 *Let $g : [a,b] \to \mathbb{R}$ be a continuous function, and let $\varphi \in C^2[\alpha, \beta]$, where the image of g is a subset of $[\alpha, \beta]$. Let $\lambda : [a,b] \to \mathbb{R}$ be a continuous function or a function of bounded variation such that $\lambda(a) \ne \lambda(b)$ and $\overline{g} \in [\alpha, \beta]$. Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ ($k = 0, 1, 2, 3, 4$) be as defined in (1)–(5). Let $p, q \in \mathbb{R}$, $1 \le p, q \le \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$.*
Then

$$\left| \frac{\int_a^b \varphi(g(x))\, d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) \right| \le Q \cdot \|\varphi''\|_p, \tag{11}$$

where

$$Q = \begin{cases} [\int_\alpha^\beta |\frac{\int_a^b G_k(g(x),s)\, d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s)|^q\, ds]^{\frac{1}{q}} & \text{for } q \ne \infty, \\ \sup_{s\in[\alpha,\beta]}\{|\frac{\int_a^b G_k(g(x),s)\, d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s)|\} & \text{for } q = \infty. \end{cases} \tag{12}$$

Proof From Lemma 1 we know that we can represent every function $\varphi \in C^2[\alpha, \beta]$ in adequate form using the previously defined functions G_k ($k = 0, 1, 2, 3, 4$). By some calculation we can easily get that, for every function $\varphi \in C^2[\alpha, \beta]$ and for any $k \in \{0, 1, 2, 3, 4\}$, we have

$$\frac{\int_a^b \varphi(g(x))\, d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) = \int_\alpha^\beta \left[\frac{\int_a^b G_k(g(x),s)\, d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s) \right] \varphi''(s)\, ds. \tag{13}$$

Taking the absolute value to (13), using the triangle inequality for integrals, and then applying the Hölder inequality, we get:

$$\left| \frac{\int_a^b \varphi(g(x))\, d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) \right|$$

$$= \left| \int_\alpha^\beta \left[\frac{\int_a^b G_k(g(x),s)\, d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s) \right] \varphi''(s)\, ds \right|$$

$$\le \int_\alpha^\beta \left| \left[\frac{\int_a^b G_k(g(x),s)\, d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s) \right] \varphi''(s) \right| ds$$

$$\leq \left(\int_{\alpha}^{\beta} \left| \frac{\int_a^b G_k(g(x),s)\,d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s) \right|^q ds \right)^{\frac{1}{q}} \cdot \left(\int_{\alpha}^{\beta} \left| \varphi''(s) \right|^p ds \right)^{\frac{1}{p}},$$

and the statement of the theorem follows. $\qquad\square$

Now consider the case $q = 1$, that is, $p = \infty$. If the positivity of the term $\frac{\int_a^b G_k(g(x),s)\,d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s)$ does not change for all $s \in [\alpha, \beta]$, then we can calculate Q. We have the following result.

Corollary 1 *Let $g : [a,b] \to \mathbb{R}$ be a continuous function, and let $\varphi \in C^2([\alpha, \beta])$, where the image of g is a subset of $[\alpha, \beta]$. Let $\lambda : [a,b] \to \mathbb{R}$ be a continuous function or a function of bounded variation such that $\lambda(a) \neq \lambda(b)$ and $\overline{g} \in [\alpha, \beta]$. Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ ($k = 0, 1, 2, 3, 4$) be as defined in* (1)–(5). *Suppose that, for any $k \in \{0, 1, 2, 3, 4\}$,*

$$\frac{\int_a^b G_k(g(x),s)\,d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s) \geq 0 \tag{14}$$

for all $s \in [\alpha, \beta]$ or that, for any $k \in \{0, 1, 2, 3, 4\}$, the reverse inequality in (14) *holds for all $s \in [\alpha, \beta]$. Then*

$$\left| \frac{\int_a^b \varphi(g(x))\,d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) \right| \leq \frac{1}{2} \cdot \left\| \varphi'' \right\|_{\infty} \cdot \left| \frac{\int_a^b (g(x))^2\,d\lambda(x)}{\int_a^b d\lambda(x)} - (\overline{g})^2 \right|. \tag{15}$$

Proof Let us start from the previous theorem and set $q = 1$, $p = \infty$. Consider any $k \in \{0, 1, 2, 3, 4\}$. Then (11) transforms into

$$\left| \frac{\int_a^b \varphi(g(x))\,d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) \right| \leq \left\| \varphi'' \right\|_{\infty} \cdot \int_{\alpha}^{\beta} \left| \frac{\int_a^b G_k(g(x),s)\,d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s) \right| ds. \tag{16}$$

When the positivity of the term $\frac{\int_a^b G_k(g(x),s)\,d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s)$ for all $s \in [\alpha, \beta]$ does not change, we can calculate the integral on the right side of (16):

$$\int_{\alpha}^{\beta} \left(\frac{\int_a^b G_k(g(x),s)\,d\lambda(x)}{\int_a^b d\lambda(x)} - G_k(\overline{g},s) \right) ds = \frac{1}{2} \cdot \left[\frac{\int_a^b (g(x))^2\,d\lambda(x)}{\int_a^b d\lambda(x)} - (\overline{g})^2 \right]. \tag{17}$$

(For the proof, see [16] and [17].)

If inequality (14) holds for all $s \in [\alpha, \beta]$, then (16) becomes

$$\left| \frac{\int_a^b \varphi(g(x))\,d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) \right| \leq \frac{1}{2} \cdot \left\| \varphi'' \right\|_{\infty} \cdot \left[\frac{\int_a^b (g(x))^2\,d\lambda(x)}{\int_a^b d\lambda(x)} - (\overline{g})^2 \right]. \tag{18}$$

Also, if the reverse inequality in (14) holds for all $s \in [\alpha, \beta]$, then (16) becomes

$$\left| \frac{\int_a^b \varphi(g(x))\,d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) \right| \leq \frac{1}{2} \cdot \left\| \varphi'' \right\|_{\infty} \cdot \left[(\overline{g})^2 - \frac{\int_a^b (g(x))^2\,d\lambda(x)}{\int_a^b d\lambda(x)} \right]. \tag{19}$$

This means that if inequality (14) or the reverse inequality in (14) holds for all $s \in [\alpha, \beta]$, then we have (15). $\qquad\square$

Remark 2 We can get the same result using the Lagrange mean-value theorems from [16] and [17], as they state that, for the functions g, φ, λ, G_k ($k = 0, 1, 2, 3, 4$) defined as in the previous theorem, if inequality (14) or the reverse inequality in (14) holds for all $s \in [\alpha, \beta]$, then there exists $\xi \in [\alpha, \beta]$ such that

$$\frac{\int_a^b \varphi(g(x)) \, d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) = \frac{1}{2} \varphi''(\xi) \left[\frac{\int_a^b (g(x))^2 \, d\lambda(x)}{\int_a^b d\lambda(x)} - (\overline{g})^2 \right]. \tag{20}$$

The next result represents an improvement of the aforementioned result from [2].

Corollary 2 *Let $g : [a, b] \to \mathbb{R}$ be continuous function, and let $\varphi \in C^2([\alpha, \beta])$, where the image of g is a subset of $[\alpha, \beta]$. Let $\lambda : [a, b] \to \mathbb{R}$ be a continuous function or a function of bounded variation such that $\lambda(a) \neq \lambda(b)$ and $\overline{g} \in [\alpha, \beta]$. Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ ($k = 0, 1, 2, 3, 4$) be as defined in (1)–(5). Let $x_0 \in [a, b]$ be arbitrarily chosen, and let $g(x_0) = c$.*

If for any $k \in \{0, 1, 2, 3, 4\}$, inequality (14) or the the reverse inequality in (14) holds for all $s \in [\alpha, \beta]$, then

$$\left| \frac{\int_a^b \varphi(g(x)) \, d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) \right| \leq \frac{1}{2} \cdot \|\varphi''\|_\infty \cdot \left[\left| \frac{\int_a^b (g(x) - c)^2 \, d\lambda(x)}{\int_a^b d\lambda(x)} \right| + \left| (\overline{g} - c)^2 \right| \right]. \tag{21}$$

Proof Let $x_0 \in [a, b]$ be arbitrarily chosen, and let $g(x_0) = c$. We have:

$$\frac{\int_a^b (g(x) - g(x_0))^2 \, d\lambda(x)}{\int_a^b d\lambda(x)} - \left(\overline{g} - g(x_0) \right)^2 = \frac{\int_a^b (g(x))^2 \, d\lambda(x)}{\int_a^b d\lambda(x)} - (\overline{g})^2. \tag{22}$$

Under the prepositions of the previous corollary, applying (22) in (15) and then using the triangle inequality, we get:

$$\left| \frac{\int_a^b \varphi(g(x)) \, d\lambda(x)}{\int_a^b d\lambda(x)} - \varphi(\overline{g}) \right| \leq \frac{1}{2} \cdot \|\varphi''\|_\infty \cdot \left| \frac{\int_a^b (g(x) - g(x_0))^2 \, d\lambda(x)}{\int_a^b d\lambda(x)} - \left(\overline{g} - g(x_0) \right)^2 \right|$$

$$\leq \frac{1}{2} \cdot \|\varphi''\|_\infty \cdot \left[\left| \frac{\int_a^b (g(x) - c)^2 \, d\lambda(x)}{\int_a^b d\lambda(x)} \right| + \left| (\overline{g} - c)^2 \right| \right]. \qquad \square$$

4 Discrete case
Discrete Jensen's inequality states that

$$\varphi \left(\frac{1}{R_n} \sum_{i=1}^n r_i x_i \right) \leq \frac{1}{R_n} \sum_{i=1}^n r_i \varphi(x_i) \tag{23}$$

for a convex function $\varphi : I \to \mathbf{R}$, $I \subseteq \mathbf{R}$, an n-tuple $\mathbf{x} = (x_1, \dots, x_n)$ ($n \geq 2$), and a nonnegative n-tuple $\mathbf{r} = (r_1, \dots, r_n)$ such that $\sum_{i=1}^n r_i > 0$.

In [16] and [17], we have a generalization of that result. It is allowed that r_i can also be negative with the sum different from 0, but there is given an additional condition on r_i, x_i in terms of the Green functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ defined in (1)–(5).

To simplify the notation, we denote $R_n = \sum_{i=1}^n r_i$ and $\overline{x} = \frac{1}{R_n} \sum_{i=1}^n r_i x_i$.

As we already know (from Lemma 1) how to represent every function $\varphi \in C^2[\alpha, \beta]$ in an adequate form using the previously defined functions G_k ($k = 0, 1, 2, 3, 4$), by some calculation it is easy to show that

$$\frac{1}{R_n} \sum_{i=1}^{n} r_i \varphi(x_i) - \varphi(\overline{x}) = \int_{\alpha}^{\beta} \left(\frac{1}{R_n} \sum_{i=1}^{n} r_i G_k(x_i, s) - G_k(\overline{x}, s) \right) \varphi''(s) \, ds. \tag{24}$$

Similarly to the integral case, applying the Hölder inequality to (24), we get the following result.

Theorem 2 *Let $x_i \in [a, b] \subseteq [\alpha, \beta]$ and $r_i \in \mathbb{R}$ ($i = 1, \ldots, n$) be such that $R_n \neq 0$ and $\overline{x} \in [\alpha, \beta]$, and let $\varphi \in C^2[\alpha, \beta]$. Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ ($k = 0, 1, 2, 3, 4$) be as defined in (1)–(5). Furthermore, let $p, q \in \mathbb{R}$, $1 \leq p, q \leq \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\left| \frac{1}{R_n} \sum_{i=1}^{n} r_i \varphi(x_i) - \varphi(\overline{x}) \right| \leq Q \cdot \| \varphi'' \|_p, \tag{25}$$

where

$$Q = \begin{cases} [\int_{\alpha}^{\beta} | \frac{1}{R_n} \sum_{i=1}^{n} r_i G_k(x_i, s) - G_k(\overline{x}, s)|^q \, ds]^{\frac{1}{q}} & \text{for } q \neq \infty; \\ \sup_{s \in [\alpha, \beta]} \{ | \frac{1}{R_n} \sum_{i=1}^{n} r_i G_k(x_i, s) - G_k(\overline{x}, s)| \} & \text{for } q = \infty. \end{cases} \tag{26}$$

Set $q = 1$, $p = \infty$. If the positivity of the term $\frac{1}{R_n} \sum_{i=1}^{n} r_i G_k(x_i, s) - G_k(\overline{x}, s)$ for all $s \in [\alpha, \beta]$ does not change, then we can calculate Q and we get the following result.

Corollary 3 *Let $x_i \in [a, b] \subseteq [\alpha, \beta]$ and $r_i \in \mathbb{R}$ ($i = 1, \ldots, n$), be such that $R_n \neq 0$ and $\overline{x} \in [\alpha, \beta]$, and let $\varphi \in C^2[\alpha, \beta]$. Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ ($k = 0, 1, 2, 3, 4$) be as defined in (1)–(5). If for any $k \in \{0, 1, 2, 3, 4\}$, the inequality*

$$\frac{1}{R_n} \sum_{i=1}^{n} r_i G_k(x_i, s) - G_k(\overline{x}, s) \geq 0 \tag{27}$$

or the reverse inequality in (27) holds for all $s \in [\alpha, \beta]$, then

$$\left| \frac{1}{R_n} \sum_{i=1}^{n} r_i \varphi(x_i) - \varphi(\overline{x}) \right| \leq \frac{1}{2} \cdot \| \varphi'' \|_\infty \cdot \left| \frac{1}{R_n} \sum_{i=1}^{n} r_i x_i^2 - \overline{x}^2 \right|. \tag{28}$$

Let $c \in [a, b] \subseteq [\alpha, \beta]$ be arbitrarily chosen. Then

$$\frac{1}{R_n} \sum_{i=1}^{n} r_i (x_i - c)^2 - (\overline{x} - c)^2 = \frac{1}{R_n} \sum_{i=1}^{n} r_i x_i^2 - \overline{x}^2, \tag{29}$$

and we have the following result.

Corollary 4 *Let $x_i \in [a, b] \subseteq [\alpha, \beta]$ and $r_i \in \mathbb{R}$ ($i = 1, \ldots, n$) be such that $R_n \neq 0$ and $\overline{x} \in [\alpha, \beta]$, and let $\varphi \in C^2[\alpha, \beta]$. Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ ($k = 0, 1, 2, 3, 4$) be as defined in (1)–(5). Let $c \in [a, b] \subseteq [\alpha, \beta]$ be arbitrarily chosen.*

If for any $k \in \{0, 1, 2, 3, 4\}$, inequality (27) *or the reverse inequality in* (27) *holds for all $s \in [\alpha, \beta]$, then*

$$\left| \frac{1}{R_n} \sum_{i=1}^n r_i \varphi(x_i) - \varphi(\overline{x}) \right| \leq \frac{1}{2} \cdot \| \varphi'' \|_\infty \cdot \left[\left| \frac{1}{R_n} \sum_{i=1}^n r_i (x_i - c)^2 \right| + \left| (\overline{x} - c)^2 \right| \right]. \tag{30}$$

5 Some applications

5.1 Applications to Csiszár f-divergence

Divergences between probability distributions have been introduced to measure the difference between them. A lot of different types of divergences exist, for example, the f-divergence, Rényi divergence, Jensen–Shannon divergence, and so on (see, e.g., [8] and [18]). There are numerous applications of divergences in many fields: anthropology and genetic, economics, ecological studies, music, signal processing, and pattern recognition.

The Jensen inequality plays an important role in obtaining inequalities for divergences between probability distributions, and there are many papers dealing with inequalities for divergences and entropies (see, e.g., [7, 9, 12]).

In this section, we give some applications of our results, and we first introduce the basic notions.

Csiszár [3, 4] defined the f-divergence functional as follows.

Definition 1 Let $f : [0, \infty\rangle \to [0, \infty\rangle$ be a convex function, and let $\mathbf{p} := (p_1, \ldots, p_n)$ and $\mathbf{q} := (q_1, \ldots, q_n)$ be positive probability distributions. The f-divergence functional is

$$D_f(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i} \right).$$

The undefined expressions can be interpreted by

$$f(0) := \lim_{t \to 0+} f(t); \qquad 0f\left(\frac{0}{0} \right) := 0; \qquad 0f\left(\frac{a}{0} \right) := \lim_{t \to 0+} tf\left(\frac{a}{t} \right), \quad a > 0.$$

This definition of the f-divergence functional can be generalized for a function $f : I \to \mathbb{R}$, $I \subset \mathbb{R}$, where $\frac{p_i}{q_i} \in I$ for $i = 1, \ldots, n$, as follows (see also [7]).

Definition 2 Let $I \subset \mathbb{R}$ be an interval, and let $f : I \to \mathbb{R}$ be a function. Let $\mathbf{p} := (p_1, \ldots, p_n) \in \mathbb{R}^n$ and $\mathbf{q} := (q_1, \ldots, q_n) \in \langle 0, \infty \rangle^n$ be such that

$$\frac{p_i}{q_i} \in I, \quad i = 1, \ldots, n.$$

Then let

$$\hat{D}_f(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i} \right).$$

Now we apply Theorem 2 to $\hat{D}_f(\mathbf{p}, \mathbf{q})$ and get the following result.

Theorem 3 *Let* $\mathbf{p} := (p_1,\ldots,p_n) \in \mathbb{R}^n$, *and* $\mathbf{q} := (q_1,\ldots,q_n) \in \langle 0,\infty \rangle^n$ *be such that*

$$\frac{p_i}{q_i} \in [a,b] \subseteq [\alpha,\beta] \quad \text{for } i = 1,\ldots,n \quad \text{and} \quad \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i} \in [\alpha,\beta].$$

Let the functions $G_k : [\alpha,\beta] \times [\alpha,\beta] \to \mathbb{R}$ $(k = 0,1,2,3,4)$ *be as defined in* (1)–(5). *Furthermore, let* $p,q \in \mathbb{R}$, $1 \le p,q \le \infty$, *be such that* $\frac{1}{p} + \frac{1}{q} = 1$.

(a) *If* $f : [\alpha,\beta] \to \mathbb{R}, f \in C^2[\alpha,\beta]$, *then*

$$\left| \frac{1}{\sum_{i=1}^n q_i} \hat{D}_f(\mathbf{p},\mathbf{q}) - f\left(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i} \right) \right| \le Q \cdot \left\| f'' \right\|_p, \tag{31}$$

where

$$Q = \begin{cases} [\int_\alpha^\beta | \frac{1}{\sum_{i=1}^n q_i} \hat{D}_{G_k(\cdot,s)}(\mathbf{p},\mathbf{q}) - G_k(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i},s)|^q \, ds]^{\frac{1}{q}} & \text{for } q \ne \infty, \\ \sup_{s \in [\alpha,\beta]} \{ | \frac{1}{\sum_{i=1}^n q_i} \hat{D}_{G_k(\cdot,s)}(\mathbf{p},\mathbf{q}) - G_k(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i},s)| \} & \text{for } q = \infty. \end{cases}$$

(b) *If* $id \cdot f \in C^2[\alpha,\beta]$, *then*

$$\left| \frac{1}{\sum_{i=1}^n q_i} \hat{D}_{id\cdot f}(\mathbf{p},\mathbf{q}) - \frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i} f\left(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i} \right) \right| \le Q \cdot \left\| (id \cdot f)'' \right\|_p, \tag{32}$$

where id *is the identity function,* $\hat{D}_{id\cdot f}(\mathbf{p},\mathbf{q}) = \sum_{i=1}^n p_i f(\frac{p_i}{q_i})$, *and*

$$Q = \begin{cases} [\int_\alpha^\beta | \frac{1}{\sum_{i=1}^n q_i} \hat{D}_{G_k(\cdot,s)}(\mathbf{p},\mathbf{q}) - G_k(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i},s)|^q \, ds]^{\frac{1}{q}} & \text{for } q \ne \infty, \\ \sup_{s \in [\alpha,\beta]} \{ | \frac{1}{\sum_{i=1}^n q_i} \hat{D}_{G_k(\cdot,s)}(\mathbf{p},\mathbf{q}) - G_k(\frac{\sum_{i=1}^n p_i}{\sum_{i=1}^n q_i},s)| \} & \text{for } q = \infty. \end{cases}$$

Proof (a) The result follows directly from Theorem 2 by substitution $\varphi := f$,

$$r_i := \frac{q_i}{\sum_{i=1}^n q_i}, \qquad x_i := \frac{p_i}{q_i}, \quad i = 1,\ldots,n.$$

(b) The result follows from (a) by substitution $f := id \cdot f$. □

Let us mention two particular cases of the previous result.

The first one corresponds to the entropy of a positive probability distribution.

Definition 3 The Shannon entropy of a positive probability distribution $\mathbf{p} := (p_1,\ldots,p_n)$ is defined by

$$H(\mathbf{p}) := -\sum_{i=1}^n p_i \log(p_i).$$

Corollary 5 *Let* $[\alpha,\beta] \subseteq \langle 0,\infty \rangle$, *and let* $\mathbf{q} := (q_1,\ldots,q_n) \in \langle 0,\infty \rangle^n$ *be such that*

$$\frac{1}{q_i} \in [a,b] \subseteq [\alpha,\beta] \quad \text{for } i = 1,\ldots,n \quad \text{and} \quad \frac{n}{\sum_{i=1}^n q_i} \in [\alpha,\beta].$$

Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ $(k = 0, 1, 2, 3, 4)$ be as defined in (1)–(5). *Furthermore, let $p, q \in \mathbb{R}$, $1 \le p, q \le \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\left| \frac{1}{\sum_{i=1}^{n} q_i} H(\mathbf{q}) - \log\left(\frac{n}{\sum_{i=1}^{n} q_i} \right) \right| \le Q \cdot \left\| \log'' \right\|_p, \tag{33}$$

where

$$Q = \begin{cases} [\int_{\alpha}^{\beta} | \frac{1}{\sum_{i=1}^{n} q_i} \sum_{i=1}^{n} q_i G_k(\frac{1}{q_i}, s) - G_k(\frac{n}{\sum_{i=1}^{n} q_i}, s)|^q \, ds]^{\frac{1}{q}} & \text{for } q \ne \infty, \\ \sup_{s \in [\alpha, \beta]} \{ | \frac{1}{\sum_{i=1}^{n} q_i} \sum_{i=1}^{n} q_i G_k(\frac{1}{q_i}, s) - G_k(\frac{n}{\sum_{i=1}^{n} q_i}, s)| \} & \text{for } q = \infty. \end{cases}$$

Proof The result follows directly from Theorem 3(a) by substitution $f := \log$ and $\mathbf{p} :=$ $(1, \ldots, 1)$.　　　　　　　　　　　　　　　　　　　　　　　　　　　　　□

The second one corresponds to the relative entropy or Kullback–Leibler divergence between two probability distributions.

Definition 4 The Kullback–Leibler divergence between two positive probability distributions $\mathbf{p} := (p_1, \ldots, p_n)$ and $\mathbf{q} := (q_1, \ldots, q_n)$ is defined by

$$D(\mathbf{p} \parallel \mathbf{q}) := \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i} \right).$$

Corollary 6 *Let $[\alpha, \beta] \subseteq \langle 0, \infty \rangle$, and let $\mathbf{p} := (p_1, \ldots, p_n) \in \mathbb{R}^n$ and $\mathbf{q} := (q_1, \ldots, q_n) \in \langle 0, \infty \rangle^n$ be such that*

$$\frac{p_i}{q_i} \in [a, b] \subseteq [\alpha, \beta] \quad \text{for } i = 1, \ldots, n \quad \text{and} \quad \frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i} \in [\alpha, \beta].$$

Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ $(k = 0, 1, 2, 3, 4)$ be as defined in (1)–(5). *Furthermore, let $p, q \in \mathbb{R}$, $1 \le p, q \le \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$.*
　　Then

$$\left| \frac{1}{\sum_{i=1}^{n} q_i} D(\mathbf{p} \parallel \mathbf{q}) - \frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i} \log\left(\frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i} \right) \right| \le Q \cdot \left\| (id \cdot \log)'' \right\|_p,$$

where id is the identity function, and

$$Q = \begin{cases} [\int_{\alpha}^{\beta} | \frac{1}{\sum_{i=1}^{n} q_i} \sum_{i=1}^{n} q_i G_k(\frac{p_i}{q_i}, s) - G_k(\frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i}, s)|^q \, ds]^{\frac{1}{q}} & \text{for } q \ne \infty, \\ \sup_{s \in [\alpha, \beta]} \{ | \frac{1}{\sum_{i=1}^{n} q_i} \sum_{i=1}^{n} q_i G_k(\frac{p_i}{q_i}, s) - G_k(\frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i}, s)| \} & \text{for } q = \infty. \end{cases}$$

Proof The result follows from Theorem 3(b) by substitution $f := \log$.　　　　　　□

5.2 Applications to Zipf–Mandelbrot law

The forthcoming results deal with the so called Zipf–Mandelbrot law.

George Kingsley Zipf (1902–1950) was a linguist who investigated the frequencies of different words in the text. The Zipf law is one of the basic laws in information science,

bibliometrics, and linguistics (see [5]). In certain fields, like economics and econometrics, this distribution is known as Pareto's law. There it analyzes the distribution of the wealthiest members of the community (see [5], p. 125). Though, in the mathematical sense, these two laws are the same; the only difference is that they are applied in different contexts (see [6], p. 294). The same kind of distribution can be also found in other scientific disciplines, such as physics, biology, earth and planetary sciences, computer science, demography, and social sciences. For more information, we refer [15].

The mathematician Benoit Mandelbrot (1924–2010) introduced a more general model of this law (see [11]). The new Zipf–Mandelbrot law has many different applications, for example, in information sciences [6], linguistics [13], ecology [14], music [10], and so on.

Definition 5 ([7]) The Zipf–Mandelbrot law is a discrete probability distribution, depending on three parameters $N \in \{1, 2, \ldots\}$, $t \in [0, \infty)$, and $\nu > 0$ and defined by

$$f(i; N, t, \nu) := \frac{1}{(i + t)^\nu H_{N,t,\nu}}, \quad i = 1, \ldots, N,$$

where

$$H_{N,t,\nu} := \sum_{j=1}^{N} \frac{1}{(j + t)^\nu}.$$

When $t = 0$, then the Zipf–Mandelbrot law becomes the Zipf law.

Now, we will apply our results for distributions given in Theorem 3 to the Zipf–Mandelbrot law, a sort of discrete probability distributions.

Corollary 7 *Let* $\mathbf{p_1}$, $\mathbf{p_2}$ *be two Zipf–Mandelbrot laws with parameters* $N \in \{1, 2, \ldots\}$, $t_1, t_2 \in [0, \infty)$, *and* $\nu_1, \nu_2 > 0$, *respectively, such that*

$$\frac{(i + t_2)^{\nu_2} H_{N,t_2,\nu_2}}{(i + t_1)^{\nu_1} H_{N,t_1,\nu_1}} \in [a, b] \subseteq [\alpha, \beta] \quad \text{for } i = 1, \ldots, N$$

$$\text{and} \quad \frac{A_1}{A_2} \in [\alpha, \beta], \quad \text{where } A_j := \sum_{i=1}^{N} \frac{1}{(i + t_j)^{\nu_j} H_{N,t_j,\nu_j}} \ (j = 1, 2).$$

Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ *(k = 0, 1, 2, 3, 4) be as defined in* (1)–(5). *Furthermore, let* $p, q \in \mathbb{R}$, $1 \le p, q \le \infty$, *be such that* $\frac{1}{p} + \frac{1}{q} = 1$.
(a) *If* $f : [\alpha, \beta] \to \mathbb{R}, f \in C^2([\alpha, \beta])$, *then*

$$\left| \frac{1}{A_2} \hat{D}_f(\mathbf{p_1}, \mathbf{p_2}) - f\left(\frac{A_1}{A_2} \right) \right| \le Q \cdot \left\| f'' \right\|_p,$$

and
(b) *if* $id \cdot f \in C^2[\alpha, \beta]$, *then*

$$\left| \frac{1}{A_2} \hat{D}_{id \cdot f}(\mathbf{p_1}, \mathbf{p_2}) - \frac{A_1}{A_2} f\left(\frac{A_1}{A_2} \right) \right| \le Q \cdot \left\| (id \cdot f)'' \right\|_p,$$

where id is the identity function, and

$$Q = \begin{cases} [\int_\alpha^\beta |\frac{1}{A_2}\hat{D}_{G_k(\cdot,s)}(\mathbf{p_1},\mathbf{p_2}) - G_k(\frac{A_1}{A_2},s)|^q \, ds]^{\frac{1}{q}} & \textit{for } q \neq \infty, \\ \sup_{s\in[\alpha,\beta]}\{|\frac{1}{A_2}\hat{D}_{G_k(\cdot,s)}(\mathbf{p_1},\mathbf{p_2}) - G_k(\frac{A_1}{A_2},s)|\} & \textit{for } q = \infty. \end{cases}$$

Although it is a particular case of the result just given, here we also present a result for the Shannon entropy.

Corollary 8 *Let $[\alpha,\beta] \subseteq \langle 0,\infty\rangle$, let $\mathbf{q}$ be the Zipf–Mandelbrot law as defined in Definition 5 such that*

$$(i+t)^\nu H_{N,t,\nu} \in [a,b] \subseteq [\alpha,\beta] \quad \textit{for } i = 1,\ldots,N$$

$$\textit{and} \quad \frac{N}{B} \in [\alpha,\beta], \quad \textit{where } B := \sum_{i=1}^N \frac{1}{(i+t)^\nu H_{N,t,\nu}}.$$

Let the functions $G_k : [\alpha,\beta] \times [\alpha,\beta] \to \mathbb{R}$ ($k = 0,1,2,3,4$) be as defined in (1)–(5). Furthermore, let $p,q \in \mathbb{R}$, $1 \leq p,q \leq \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$.
Then

$$\left| \frac{1}{B}H(\mathbf{q}) - \log\left(\frac{N}{B}\right) \right| \leq Q \cdot \|\log''\|_p,$$

where

$$Q = \begin{cases} [\int_\alpha^\beta |\frac{1}{B}\sum_{i=1}^N \frac{1}{(i+t)^\nu H_{N,t,\nu}} G_k((i+t)^\nu H_{N,t,\nu};s) - G_k(\frac{N}{B},s)|^q \, ds]^{\frac{1}{q}} & \textit{for } q \neq \infty, \\ \sup_{s\in[\alpha,\beta]}\{|\frac{1}{B}\sum_{i=1}^N \frac{1}{(i+t)^\nu H_{N,t,\nu}} G_k((i+t)^\nu H_{N,t,\nu};s) - G_k(\frac{N}{B},s)|\} & \textit{for } q = \infty. \end{cases}$$

Author details
[1]Catholic University of Croatia, Zagreb, Croatia. [2]Faculty of Textile Technology, University of Zagreb, Zagreb, Croatia. [3]RUDN University, Moscow, Russia.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Agarwal, R.P., Wong, P.J.Y.: Error Inequalities in Polynomial Interpolation and Their Applications. Kluwer Academic, Dordrecht (1993)
2. Costarelli, D., Spigler, R.: How sharp is the Jensen inequality? J. Inequal. Appl. **2015**, 69 (2015). https://doi.org/10.1186/s13660-015-0591-x
3. Csiszár, I.: Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hung. **2**, 299–318 (1967)
4. Csiszár, I.: Information measures: a critical survey. In: Trans. 7th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Processes and 8th European Meeting of Statist., vol. B, pp. 73–86. Academia, Prague (1978)
5. Diodato, V.: Dictionary of Bibliometrics. Haworth Press, New York (1994)
6. Egghe, L., Rousseau, R.: Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science. Elsevier, New York (1990)
7. Horváth, L., Pečarić, Đ., Pečarić, J.: Estimations of f- and Rényi divergences by using a cyclic refinement of the Jensen's inequality. Bull. Malays. Math. Sci. Soc. (2017). https://doi.org/10.1007/s40840-017-0526-4
8. Liese, F., Vajda, I.: Convex Statistical Distances. Teubner-Texte Zur Mathematik, vol. 95. Teubner, Leipzig (1987)
9. Lovričević, N., Pečarić, Đ., Pečarić, J.: Zipf–Mandelbrot law, f-divergences and the Jensen-type interpolating inequalities. J. Inequal. Appl. **2018**, 36 (2018). https://doi.org/10.1186/s13660-018-1625-y
10. Manaris, B., Vaughan, D., Wagner, C.S., Romero, J., Davis, R.B.: Evolutionary music and the Zipf–Mandelbrot law: developing fitness functions for pleasant music. In: Proceedings of 1st European Workshop on Evolutionary Music and Art (EvoMUSART2003), pp. 522–534 (2003)
11. Mandelbrot, B.: An informational theory of the statistical structure of language. In: Jackson, W. (ed.) Communication Theory, pp. 486–502. Academic Press, New York (1953)
12. Mikić, R., Pečarić, Đ., Pečarić, J.: Inequalities of the Jensen and Edmundson–Lah–Ribarič type for 3-convex functions with applications. J. Math. Inequal. **12**(3), 677–692 (2018). https://doi.org/10.7153/jmi-2018-12-52
13. Montemurro, M.A.: Beyond the Zipf–Mandelbrot law in quantitative linguistics. arXiv:cond-mat/0104066v2 (2001)
14. Mouillot, D., Lepretre, A.: Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity. In: Environmental Monitoring and Assessment, vol. 63, pp. 279–295. Springer, Berlin (2000)
15. Newman, M.E.J.: Power laws, Pareto distributions and Zipf's law. http://arxiv.org/pdf/cond-mat/0412004.pdf?origin=publication_detail (30.9.2018)
16. Pečarić, J., Perić, I., Rodić Lipanović, M.: Uniform treatment of Jensen type inequalities. Math. Rep. **16**(66), 2, 183–205 (2014)
17. Pečarić, J., Rodić, M.: Uniform treatment of Jensen type inequalities II. Math. Rep. Accepted
18. Vajda, I.: Theory of Statistical Inference and Information. Kluwer Academic, Dordrecht (1989)