Mathematical Inequalities & Applications
Volume 22, Number 4 (2019), 1463–1475 doi:10.7153/mia-2019-22-102

ON A JENSEN–TYPE INEQUALITY FOR GENERALIZED f–DIVERGENCES AND ZIPF–MANDELBROT LAW

ĐILDA PEČARIĆ, JOSIP PEČARIĆ AND MIRNA RODIĆ*

(*Communicated by M. Praljak*)

Abstract. By means of one new Jensen-type inequality for signed measures which is characterized via several different Green functions, in this paper we derive new inequalities for generalized $f-$ divergences. The applications on the Zipf-Mandelbrot law, as one specific kind of probability distributions, are also given.

1. Introduction

The divergences measure the differences between probability distributions. They are applied in many different fields like economics, ecology, biology, genetics, anthropology, information theory, signal processing, etc. Different authors investigated these measures of difference and defined different types of divergences. So we can read about f-divergence, Rényi divergence, Jensen-Shannon divergence, χ^{α}-divergences, elementary divergences, Matusita's divergences, Puri-Vincze divergences, divergences of Arimoto-type, perimeter-type divergences, etc. (An interested reader can also consult [6], [9] and [16]).

Jensen's inequality is important in obtaining inequalities for divergences between probability distributions, and there are many papers dealing with inequalities for divergences and entropies (see for example [4] or [10]).

By means of one Jensen-type inequality for signed measures which is characterized via several different Green functions, in this paper we will derive some new inequalities for divergences. At the end, we will also give the applications on the Zipf-Mandelbrot law, as one specific kind of probability distributions.

2. Preliminary results

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

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

Mathematics subject classification (2010): 26D15, 60E05, 94A17.
Keywords and phrases: Green function, Jensen inequality, $f-$divergence, Zipf-Mandelbrot law.
* Corresponding author.

Paper MIA-22-102

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

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

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

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

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

The following lemma holds (see [13] and [14]):

LEMMA 1. *For every function $\varphi:[\alpha,\beta]\to\mathbb{R},\ \varphi\in C^2([\alpha,\beta])$, the following identities hold:*

$$\begin{aligned}
\varphi(x)&=\frac{\beta-x}{\beta-\alpha}\varphi(\alpha)+\frac{x-\alpha}{\beta-\alpha}\varphi(\beta)+\int_\alpha^\beta G_0(x,s)\varphi''(s)ds,\\
\varphi(x)&=\varphi(\alpha)+(x-\alpha)\varphi'(\beta)+\int_\alpha^\beta G_1(x,s)\varphi''(s)ds,\\
\varphi(x)&=\varphi(\beta)+(x-\beta)\varphi'(\alpha)+\int_\alpha^\beta G_2(x,s)\varphi''(s)ds,\\
\varphi(x)&=\varphi(\beta)-(\beta-\alpha)\varphi'(\beta)+(x-\alpha)\varphi'(\alpha)+\int_\alpha^\beta G_3(x,s)\varphi''(s)ds,\\
\varphi(x)&=\varphi(\alpha)+(\beta-\alpha)\varphi'(\alpha)-(\beta-x)\varphi'(\beta)+\int_\alpha^\beta G_4(x,s)\varphi''(s)ds,
\end{aligned}$$

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

This lemma was crucial in establishing the uniform treatment of the Jensen-type inequalities, giving the necessary and sufficient conditions for such inequalities to hold in case of the not necessarily positive real Stieltjes measure (see [13] and [14]).

As we are interested in probability distributions here, in this paper we will consider the discrete results.

3. Discrete Jensen-type inequality

The discrete Jensen inequality states that

$$\varphi\left(\frac{1}{U_n}\sum_{i=1}^n u_i x_i\right)\leqslant\frac{1}{U_n}\sum_{i=1}^n u_i\varphi(x_i)$$

holds for a convex function $\varphi : I \to \mathbf{R}$, $I \subseteq \mathbf{R}$, an n-tuple $\mathbf{x} = (x_1, \ldots, x_n)$ $(n \geqslant 2)$ and nonnegative n-tuple $\mathbf{u} = (u_1, \ldots, u_n)$, such that $U_n = \sum_{i=1}^{n} u_i > 0$. In [13] and [14] we have the generalization of that result. Namely, there is also allowed that u_i are negative with their sum different from 0, but we have a supplementary demand on u_i, x_i using the Green functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ defined in (1)-(5).

In order to simplify the notation in this paper we shall use the common notation: $U_n = \sum_{i=1}^{n} u_i$ and and $\overline{x} = \frac{1}{U_n} \sum_{i=1}^{n} u_i x_i$.

The following result holds true:

THEOREM 1. *Let $x_i \in [a,b] \subseteq [\alpha, \beta]$, $u_i \in \mathbb{R}$ $(i = 1, \ldots, n)$, be such that $U_n \neq 0$ and $\overline{x} \in [\alpha, \beta]$, and let $\varphi : [\alpha, \beta] \to \mathbb{R}$, $\varphi \in C^2([\alpha, \beta])$. 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 $p, q \in \mathbb{R}$, $1 \leqslant p, q \leqslant \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\left| \frac{1}{U_n} \sum_{i=1}^{n} u_i \varphi(x_i) - \varphi(\overline{x}) \right| \leqslant Q \cdot \left\| \varphi'' \right\|_p$$

holds, where

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

Proof. As we already know (from Lemma 1) how to represent every function $\varphi : [\alpha, \beta] \to \mathbb{R}$, $\varphi \in C^2([\alpha, \beta])$, in adequate form using previously defined functions G_k $(k = 0, 1, 2, 3, 4)$, it's easy to show by some calculation that for every such function φ and for any $k \in \{0, 1, 2, 3, 4\}$ it holds:

$$\frac{1}{U_n} \sum_{i=1}^{n} u_i \varphi(x_i) - \varphi(\overline{x}) = \int_\alpha^\beta \left(\frac{1}{U_n} \sum_{i=1}^{n} u_i G_k(x_i, s) - G_k(\overline{x}, s) \right) \varphi''(s) ds. \tag{6}$$

Applying the absolute value on (6), and using the triangle inequality for integrals which says that for every function f the following is valid

$$\left| \int_a^b f(x) dx \right| \leqslant \int_a^b |f(x)| \, dx,$$

applying the Hölder inequality we get the following:

$$\left| \frac{1}{U_n} \sum_{i=1}^{n} u_i \varphi(x_i) - \varphi(\overline{x}) \right| = \left| \int_\alpha^\beta \left(\frac{1}{U_n} \sum_{i=1}^{n} u_i G_k(x_i, s) - G_k(\overline{x}, s) \right) \varphi''(s) ds \right|$$

$$\leqslant \int_\alpha^\beta \left| \left(\frac{1}{U_n} \sum_{i=1}^{n} u_i G_k(x_i, s) - G_k(\overline{x}, s) \right) \varphi''(s) \right| ds$$

$$\leqslant \left(\int_\alpha^\beta \left| \frac{1}{U_n} \sum_{i=1}^{n} u_i G_k(x_i, s) - G_k(\overline{x}, s) \right|^q ds \right)^{\frac{1}{q}} \cdot \left(\int_\alpha^\beta \left| \varphi''(s) \right|^p ds \right)^{\frac{1}{p}},$$

and we get the result given in our theorem. □

REMARK 1. The analogue of the previous theorem for the integral case can be found in [12].

4. Inequalities for different types of generalized $f-$divergences

For a function $f : \mathbb{R}_+ \to \mathbb{R}$ and positive probability distributions $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$, I. Csiszár (in [1], [2]) defined the f-divergence functional by

$$C_f(\mathbf{q},\mathbf{p}) := \sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right). \tag{7}$$

The undefined expressions can be interpreted as follows

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

I. Csiszár studied (7) under assumption that function f is convex. Independently, some other authors also introduced and studied these divergences, but (7) is widely known as the *Csiszár $f-$divergence*.

The definition of the f-divergence functional can be further generalized, and we have the following generalization given in [11] which uses weights.

For a function $f : \mathbb{R}_+ \to \mathbb{R}$ and $\mathbf{p} = (p_1, \ldots, p_n) \in \mathbb{R}^n_+$, $\mathbf{q} = (q_1, \ldots, q_n) \in \mathbb{R}^n_+$, $\mathbf{r} := (r_1, \ldots, r_n) \in \mathbb{R}^n_+$, the generalized Csiszár $f-$divergence is defined by ([11])

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) := \sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right). \tag{8}$$

We can now apply Theorem 1 on $C_f(\mathbf{q},\mathbf{p};\mathbf{r})$, and we get our next result.

In order to simplify our results, we introduce the following notation:

$$P_r = \sum_{i=1}^{n} r_i p_i, \tag{9}$$

$$\overline{Q}_r = \frac{1}{P_r}\sum_{i=1}^{n} r_i q_i. \tag{10}$$

THEOREM 2. *Let $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathbb{R}^n_+$ be such that*

$$\frac{q_i}{p_i} \in [a,b] \subseteq [\alpha,\beta] \textit{ for } i = 1,\ldots,n; \textit{ and that } \overline{Q}_r \in [\alpha,\beta],$$

where $\overline{Q}_r$ is as defined in (10).

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 $p,q \in \mathbb{R}$, $1 \leqslant p,q \leqslant \infty$, 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*

$$\left| \frac{1}{P_r} C_f(\mathbf{q},\mathbf{p};\mathbf{r}) - f\left(\overline{Q}_r\right) \right| \leqslant Q \cdot \left\| f'' \right\|_p$$

holds, where P_r *and* $C_f(\mathbf{q},\mathbf{p};\mathbf{r})$ *are as defined in (9) and (8) respectively, and*

$$Q = \begin{cases} \left[\int_\alpha^\beta \left| \frac{1}{P_r} \sum_{i=1}^n r_i p_i G_k\left(\frac{q_i}{p_i}, s\right) - G_k\left(\overline{Q}_r, s\right) \right|^q ds \right]^{\frac{1}{q}}, & \text{for } q \neq \infty; \\ \sup_{s\in[\alpha,\beta]} \left\{ \left| \frac{1}{P_r} \sum_{i=1}^n r_i p_i G_k\left(\frac{q_i}{p_i}, s\right) - G_k\left(\overline{Q}_r, s\right) \right| \right\}, & \text{for } q = \infty. \end{cases} \tag{11}$$

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

$$\left| \frac{1}{P_r} C_{id\cdot f}(\mathbf{q},\mathbf{p};\mathbf{r}) - \overline{Q}_r \cdot f\left(\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i} \right) \right| \leqslant Q \cdot \left\| (id \cdot f)'' \right\|_p$$

holds, where id *is the identity function,* $C_{id\cdot f}(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^n r_i q_i f\left(\frac{q_i}{p_i}\right)$ *and* Q *is as defined in (11).*

Proof.

(a) The result follows directly from Theorem 1 by substitution $\varphi := f$,

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

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

In the following results we consider some of the most important examples of $f-$divergences.

For $\mathbf{p},\mathbf{q} \in \mathbb{R}_+^n$, the *Kullback-Leibler divergence* is defined by (see [7], [8])

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

It is easy to see that the Kullback-Leibler divergence is in fact Csiszár f-divergence if we set $f(t) = t\log t,\ t > 0$.

The generalized Kullback-Leibler divergence is defined by (see [11])

$$KL(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^n r_i q_i \log\frac{q_i}{p_i},$$

where $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathbb{R}_+^n$.

PROPOSITION 1. *Let* $\mathbf{p},\mathbf{q},\mathbf{r}\in\mathbb{R}_+^n$ *be such that*

$$\frac{q_i}{p_i}\in[a,b]\subseteq[\alpha,\beta]\text{ for }i=1,\dots,n;\text{ and that }\overline{Q}_r\in[\alpha,\beta],$$

where $\overline{Q}_r$ *is as defined in (10).*

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* $p,q\in\mathbb{R}$, $1\leqslant p,q\leqslant\infty$, *be such that* $\frac{1}{p}+\frac{1}{q}=1$.

Then

$$\left|\frac{1}{P_r}KL(\mathbf{q},\mathbf{p};\mathbf{r})-\overline{Q}_r\cdot\log\left(\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i}\right)\right|\leqslant Q\cdot\left\|(id\cdot\log)''\right\|_p$$

holds, where P_r *is as defined in (9),* id *is the identity function and* Q *is as defined in (11).*

Proof. The result follows from Theorem 2 (b) by substitution $f:=\log$ (i.e. from Theorem 2 (a) by substitution $f(t):=t\log(t)$, $t>0$). $\square$

For $\mathbf{p},\mathbf{q}\in\mathbb{R}_+^n$, the *Hellinger divergence* is defined by (see [3])

$$He(\mathbf{q},\mathbf{p})=\sum_{i=1}^n(\sqrt{q_i}-\sqrt{p_i})^2.$$

The Hellinger divergence is also the Csiszár f-divergence where $f(t)=(1-\sqrt{t})^2$, $t>0$.

The generalization of the Hellinger divergence for $\mathbf{r}\in\mathbb{R}_+^n$ is defined by ([11])

$$He(\mathbf{q},\mathbf{p};\mathbf{r})=\sum_{i=1}^n r_i(\sqrt{q_i}-\sqrt{p_i})^2.$$

PROPOSITION 2. *Let* $\mathbf{p},\mathbf{q},\mathbf{r}\in\mathbb{R}_+^n$ *be such that*

$$\frac{q_i}{p_i}\in[a,b]\subseteq[\alpha,\beta]\text{ for }i=1,\dots,n;\text{ and that }\overline{Q}_r\in[\alpha,\beta],$$

where $\overline{Q}_r$ *is as defined in (10).*

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* $p,q\in\mathbb{R}$, $1\leqslant p,q\leqslant\infty$, *be such that* $\frac{1}{p}+\frac{1}{q}=1$.

Then

$$\left|\frac{1}{P_r}He(\mathbf{q},\mathbf{p};\mathbf{r})-\left(1-\sqrt{\overline{Q}_r}\right)^2\right|\leqslant Q\cdot\left\|f''\right\|_p$$

holds, where P_r *is as defined in (9),* $f(t)=(1-\sqrt{t})^2$, $t>0$, *and* Q *is as defined in (11).*

Proof. For $f(t) = (1-\sqrt{t})^2$, $t > 0$, we have that

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i \left(1-\sqrt{\frac{q_i}{p_i}}\right)^2 = \sum_{i=1}^{n} r_i(\sqrt{p_i}-\sqrt{q_i})^2 = \sum_{i=1}^{n} r_i(\sqrt{q_i}-\sqrt{p_i})^2 = He(\mathbf{q},\mathbf{p};\mathbf{r}),$$

and the statement from our proposition follows from Theorem 2 (a). $\square$

For $\mathbf{p},\mathbf{q} \in \mathbb{R}_+^n$, the *Rényi divergence* is defined by ([15])

$$Re_\gamma(\mathbf{q},\mathbf{p}) = \sum_{i=1}^{n} q_i^\gamma p_i^{1-\gamma}, \ \gamma \in \langle 1, +\infty \rangle.$$

Its generalization for $\mathbf{r} \in \mathbb{R}_+^n$ is defined by ([11])

$$Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i q_i^\gamma p_i^{1-\gamma}.$$

We have the following result.

PROPOSITION 3. *Let $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathbb{R}_+^n$ be such that*

$$\frac{q_i}{p_i} \in [a,b] \subseteq [\alpha,\beta] \ \text{for } i = 1,\dots,n; \ \text{and that } \overline{Q}_r \in [\alpha,\beta],$$

where $\overline{Q}_r$ is as defined in (10).

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 $p,q \in \mathbb{R}$, $1 \leqslant p,q \leqslant \infty$, be such that $\frac{1}{p}+\frac{1}{q} = 1$.

Then

$$\left|\frac{1}{P_r} Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r}) - \overline{Q}_r^\gamma\right| \leqslant Q \cdot \left\|f''\right\|_p$$

holds, where P_r is as defined in (9), $f(t) = t^\gamma$ ($t > 0$, $\gamma > 1$), and Q is as defined in (11).

Proof. The result follows from Theorem 2 (a), as for $f(t) = t^\gamma$ $(t > 0,\ \gamma > 1)$ it holds

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i \left(\frac{q_i}{p_i}\right)^\gamma = \sum_{i=1}^{n} r_i q_i^\gamma p_i^{1-\gamma} = Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r}). \quad \square$$

For $\mathbf{p},\mathbf{q} \in \mathbb{R}_+^n$, the χ^2-*divergence* is defined by

$$D_{\chi^2}(\mathbf{q},\mathbf{p}) = \sum_{i=1}^{n} \frac{(q_i-p_i)^2}{p_i}.$$

The generalized χ^2-divergence for $\mathbf{r} \in \mathbb{R}_+^n$ is defined by ([11])

$$D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i \frac{(q_i-p_i)^2}{p_i}.$$

The following result holds.

PROPOSITION 4. *Let* $\mathbf{p}, \mathbf{q}, \mathbf{r} \in \mathbb{R}_+^n$ *be such that*

$$\frac{q_i}{p_i} \in [a,b] \subseteq [\alpha,\beta] \text{ for } i = 1,\dots,n; \text{ and that } \overline{Q}_r \in [\alpha,\beta],$$

where $\overline{Q}_r$ *is as defined in (10).*

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* $p, q \in \mathbb{R}$, $1 \leqslant p, q \leqslant \infty$, *be such that* $\frac{1}{p} + \frac{1}{q} = 1$.

Then

$$\left| \frac{1}{P_r} D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}) - (\overline{Q}_r - 1)^2 \right| \leqslant Q \cdot \left\| f'' \right\|_p$$

holds, where P_r *is as defined in (9),* $f(t) = (t-1)^2$, $t > 0$, *and* Q *is as defined in (11).*

Proof. For $f(t) = (t-1)^2$, $t > 0$, we have that

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^n r_i p_i \left(\frac{q_i}{p_i} - 1 \right)^2 = \sum_{i=1}^n r_i \frac{(q_i - p_i)^2}{p_i} = D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}),$$

so the statement from our proposition follows from Theorem 2 (a). □

The Shannon entropy of a positive probability distribution $\mathbf{p} = (p_1,\dots,p_n)$ is defined by (see [4])

$$H(\mathbf{p}) = -\sum_{i=1}^n p_i \log(p_i). \tag{12}$$

We can see that (12) is a special case of the Csiszár f-divergence $C_f(\mathbf{q},\mathbf{p})$ if we set $\mathbf{q} = (1,\dots,1) \in \mathbb{R}_+^n$ and function $f(t) = \log t$, $t > 0$. We can also consider the generalized Shannon entropy which is defined by

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

We have the following result.

PROPOSITION 5. *Let* $\mathbf{p}, \mathbf{r} \in \mathbb{R}_+^n$ *be such that*

$$\frac{1}{p_i} \in [a,b] \subseteq [\alpha,\beta] \text{ for } i = 1,\dots,n; \text{ and that } \frac{1}{P_r} \sum_{i=1}^n r_i \in [\alpha,\beta]$$

where P_r *is as defined in (9).*

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* $p, q \in \mathbb{R}$, $1 \leqslant p, q \leqslant \infty$, *be such that* $\frac{1}{p} + \frac{1}{q} = 1$.

Then

$$\left| \frac{1}{P_r} H(\mathbf{p};\mathbf{r}) - \log\left(\frac{1}{P_r} \sum_{i=1}^n r_i \right) \right| \leqslant Q \cdot \left\| \log'' \right\|_p$$

holds, where Q *is as defined in (11).*

Proof. The result follows from Theorem 2 (a) by substitution $f := \log$ and $\mathbf{q} = (1,\dots,1)$. □

5. Applications to Zipf-Mandelbrot law

DEFINITION 1. [4] (see also [5]) Zipf-Mandelbrot law is a discrete probability distribution, depends on three parameters $N \in \{1,2,\dots\}$, $t \in [0,\infty\rangle$ and $v > 0$, and it is defined by

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

where

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

When $t = 0$, then Zipf–Mandelbrot law becomes Zipf's law.

Now, we can apply our results for distributions on the Zipf-Mandelbrot law, as a sort of discrete probability distribution.

Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\dots\}$, t_1, $t_2 \geqslant 0$ and v_1, $v_2 > 0$, respectively. It is

$$p_i = f(i;N,t_1,v_1) := \frac{1}{(i+t_1)^{v_1} H_{N,t_1,v_1}}, \quad i = 1,\dots,N, \tag{13}$$

and

$$q_i = f(i;N,t_2,v_2) := \frac{1}{(i+t_2)^{v_2} H_{N,t_2,v_2}}, \quad i = 1,\dots,N, \tag{14}$$

where

$$H_{N,t_k,v_k} := \sum_{j=1}^{N} \frac{1}{(j+t_k)^{v_k}}, \quad k = 1,2.$$

Then the generalized Csiszár divergence for such $\mathbf{p},\mathbf{q}$, and for $\mathbf{r} \in \mathbb{R}_+^n$ is given by

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \frac{1}{H_{N,t_1,v_1}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} f\left(\frac{(i+t_1)^{v_1} H_{N,t_1,v_1}}{(i+t_2)^{v_2} H_{N,t_2,v_2}}\right). \tag{15}$$

Using (13) and (14), we have the following expressions for (9) and (10)

$$P_r = \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1} H_{N,t_1,v_1}} = \frac{1}{H_{N,t_1,v_1}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}, \tag{16}$$

$$\overline{Q}_r = \frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2} H_{N,t_2,v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1} H_{N,t_1,v_1}}} = \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} \cdot \frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}}. \tag{17}$$

We have the following result.

COROLLARY 1. *Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\dots\}$, t_1, $t_2 \geqslant 0$ and $v_1, v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}_+^n$ such that*

$$\frac{q_i}{p_i} := \frac{(i+t_1)^{v_1} H_{N,t_1,v_1}}{(i+t_2)^{v_2} H_{N,t_2,v_2}} \in [a,b] \subseteq [\alpha,\beta] \; \textit{for } i = 1,\dots,N,$$

and that $\overline{Q}_r \in [\alpha,\beta]$, *where* $\overline{Q}_r$ *is as defined in* (17).

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* $p,q \in \mathbb{R}$, $1 \leqslant p,q \leqslant \infty$, *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*

$$\left|\frac{1}{P_r}C_f(\mathbf{q},\mathbf{p};\mathbf{r}) - f\left(\overline{Q}_r\right)\right| \leqslant Q\cdot\left\|f''\right\|_p$$

holds, and

(b) *if* $id\cdot f : [\alpha,\beta] \to \mathbb{R}$, $id\cdot f \in C^2([\alpha,\beta])$, *then*

$$\left|\frac{1}{P_r}C_{id\cdot f}(\mathbf{q},\mathbf{p};\mathbf{r}) - \overline{Q}_r\cdot f\left(\frac{\sum_{i=1}^N q_i}{\sum_{i=1}^N p_i}\right)\right| \leqslant Q\cdot\left\|(id\cdot f)''\right\|_p$$

holds, where id *is the identity function,* Q, p_i, q_i, P_r, $C_f(\mathbf{q},\mathbf{p};\mathbf{r})$ *are as defined in* (11), (13), (14), (16), (15) *respectively.*

If $\mathbf{p},\mathbf{q}$ are two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geqslant 0$ and v_1, $v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}_+^n$, for the generalized Kullbach-Leibler divergence we have the following representation:

$$KL(\mathbf{q},\mathbf{p};\mathbf{r}) = \frac{1}{H_{N,t_2,v_2}}\sum_{i=1}^N \frac{r_i}{(i+t_2)^{v_2}}\log\left(\frac{(i+t_1)^{v_1}H_{N,t_1,v_1}}{(i+t_2)^{v_2}H_{N,t_2,v_2}}\right). \tag{18}$$

The following result holds true:

COROLLARY 2. *Let* $\mathbf{p},\mathbf{q}$ *be two Zipf-Mandelbrot laws with parameters* $N \in \{1,2,\ldots\}$, t_1, $t_2 \geqslant 0$ *and* $v_1, v_2 > 0$, *respectively, and* $\mathbf{r} \in \mathbb{R}_+^n$ *such that*

$$\frac{q_i}{p_i} := \frac{(i+t_1)^{v_1}H_{N,t_1,v_1}}{(i+t_2)^{v_2}H_{N,t_2,v_2}} \in [a,b] \subseteq [\alpha,\beta] \text{ for } i = 1,\ldots,N,$$

and that $\overline{Q}_r \in [\alpha,\beta]$, *where* $\overline{Q}_r$ *is as defined in* (17).

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* $p,q \in \mathbb{R}$, $1 \leqslant p,q \leqslant \infty$, *be such that* $\frac{1}{p}+\frac{1}{q} = 1$.

Then

$$\left|\frac{1}{P_r}KL(\mathbf{q},\mathbf{p};\mathbf{r}) - \overline{Q}_r\cdot\log\left(\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i}\right)\right| \leqslant Q\cdot\left\|(id\cdot\log)''\right\|_p$$

holds, where id *is the identity function,* Q, p_i, q_i, P_r, $KL(\mathbf{q},\mathbf{p};\mathbf{r})$ *are as defined in* (11), (13), (14), (16), (18) *respectively.*

For $\mathbf{p},\mathbf{q}$ two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geqslant 0$ and v_1, $v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}_+^n$, the generalized Hellinger divergence has the following representation:

$$He(\mathbf{q},\mathbf{p};\mathbf{r}) = \frac{1}{H_{N,t_1,v_1} H_{N,t_2,v_2}} \sum_{i=1}^{N} r_i \frac{\left(\sqrt{(i+t_1)^{v_1} H_{N,t_1,v_1}} - \sqrt{(i+t_2)^{v_2} H_{N,t_2,v_2}}\right)^2}{(i+t_1)^{v_1}(i+t_2)^{v_2}}. \tag{19}$$

The following result holds true:

COROLLARY 3. *Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geqslant 0$ and $v_1, v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}_+^n$ such that*

$$\frac{q_i}{p_i} := \frac{(i+t_1)^{v_1} H_{N,t_1,v_1}}{(i+t_2)^{v_2} H_{N,t_2,v_2}} \in [a,b] \subseteq [\alpha,\beta] \text{ for } i = 1,\ldots,N,$$

and that $\overline{Q}_r \in [\alpha,\beta]$, where $\overline{Q}_r$ is as defined in (17).

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 $p,q \in \mathbb{R}$, $1 \leqslant p,q \leqslant \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$.

Then

$$\left| \frac{1}{P_r} He(\mathbf{q},\mathbf{p};\mathbf{r}) - \left(1 - \sqrt{\overline{Q}_r}\right)^2 \right| \leqslant Q \cdot \left\| f'' \right\|_p$$

holds, where Q, p_i, q_i, P_r, $He(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (11), (13), (14), (16), (19) respectively, and $f(t) = (1-\sqrt{t})^2$, $t > 0$.

For $\mathbf{p},\mathbf{q}$ two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geqslant 0$ and v_1, $v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}_+^n$, the generalized Rényi divergence has the following representation:

$$Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r}) = \frac{H_{N,t_1,v_1}^{\gamma-1}}{H_{N,t_2,v_2}^{\gamma}} \sum_{i=1}^{N} r_i \frac{(i+t_1)^{(\gamma-1)v_1}}{(i+t_2)^{\gamma v_2}}. \tag{20}$$

The following result holds true:

COROLLARY 4. *Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geqslant 0$ and $v_1, v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}_+^n$ such that*

$$\frac{q_i}{p_i} := \frac{(i+t_1)^{v_1} H_{N,t_1,v_1}}{(i+t_2)^{v_2} H_{N,t_2,v_2}} \in [a,b] \subseteq [\alpha,\beta] \text{ for } i = 1,\ldots,N,$$

and that $\overline{Q}_r \in [\alpha,\beta]$, where $\overline{Q}_r$ is as defined in (17).

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 $p,q \in \mathbb{R}$, $1 \leqslant p,q \leqslant \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$.

Then

$$\left| \frac{1}{P_r} Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r}) - \overline{Q}_r^{\gamma} \right| \leqslant Q \cdot \left\| f'' \right\|_p$$

holds, where Q, p_i, q_i, P_r, $Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (11), (13), (14), (16), (20) respectively, and $f(t) = t^\gamma$, $(t > 0, \gamma > 1)$.

For $\mathbf{p}, \mathbf{q}$ two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geqslant 0$ and v_1, $v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}_+^n$, the generalized χ^2-divergence has the following representation:

$$D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}) = H_{N,t_1,v_1} \cdot \sum_{i=1}^{N} r_i (i+t_1)^{v_1} \left(\frac{1}{(i+t_2)^{v_2} H_{N,t_2,v_2}} - \frac{1}{(i+t_1)^{v_1} H_{N,t_1,v_1}} \right)^2. \quad (21)$$

The following result holds true:

COROLLARY 5. *Let $\mathbf{p}, \mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geqslant 0$ and $v_1, v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}_+^n$ such that*

$$\frac{q_i}{p_i} := \frac{(i+t_1)^{v_1} H_{N,t_1,v_1}}{(i+t_2)^{v_2} H_{N,t_2,v_2}} \in [a,b] \subseteq [\alpha,\beta] \text{ for } i = 1,\ldots,N,$$

and that $\overline{Q}_r \in [\alpha,\beta]$, where $\overline{Q}_r$ is as defined in (17).

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 $p,q \in \mathbb{R}$, $1 \leqslant p,q \leqslant \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$.

Then

$$\left| \frac{1}{P_r} D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}) - (\overline{Q}_r - 1)^2 \right| \leqslant Q \cdot \left\| f'' \right\|_p$$

holds, where Q, p_i, q_i, P_r, $D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (11), (13), (14), (16), (21) respectively, and $f(t) = (t-1)^2$, $t > 0$.

If $\mathbf{p}$ is the Zipf-Mandelbrot law with parameters $N \in \{1,2,\ldots\}$, $t_1 \geqslant 0$ and $v_1 > 0$, and $\mathbf{r} \in \mathbb{R}_+^n$, then the generalized Shannon entropy $H(\mathbf{p};\mathbf{r})$ has the following representation:

$$H(\mathbf{p};\mathbf{r}) = \frac{1}{H_{N,t_1,v_1}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} \log\left[(i+t_1)^{v_1} H_{N,t_1,v_1}\right]. \quad (22)$$

We have the following result.

COROLLARY 6. *Let $\mathbf{p}$ be the Zipf-Mandelbrot law with parameters $N \in \{1,2,\ldots\}$, $t_1 \geqslant 0$ and $v_1 > 0$, and $\mathbf{r} \in \mathbb{R}_+^n$ such that*

$$\frac{1}{p_i} := (i+t_1)^{v_1} H_{N,t_1,v_1} \in [a,b] \subseteq [\alpha,\beta] \text{ for } i = 1,\ldots,N,$$

and that $\frac{1}{P_r} \sum_{i=1}^{n} r_i \in [\alpha,\beta]$, where P_r is as defined in (16).

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 $p,q \in \mathbb{R}$, $1 \leqslant p,q \leqslant \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$.

Then

$$\left| \frac{1}{P_r} H(\mathbf{p};\mathbf{r}) - \log\left(\frac{1}{P_r} \sum_{i=1}^{n} r_i \right) \right| \leqslant Q \cdot \left\| \log'' \right\|_p$$

holds, where Q, p_i, q_i, $H(\mathbf{p};\mathbf{r})$ are as defined in (11), (13), (14), (22) respectively.

REFERENCES

[1] I. CSISZÁR, *Information measures: A critical survey*, Trans. 7th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Processes and 8th European Meeting of Statist., Volume B, 73–86, Academia Prague 1978.

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

[3] S. S. DRAGOMIR, *Some inequalities for the Csiszár ϕ-divergence when ϕ is an L-Lipschitzian function and applications*, Ital. J. Pure Appl. Math., 15, 2004, 57–76.

[4] L. HORVÁTH, Đ. PEČARIĆ AND J. PEČARIĆ, *Estimations of $f-$ and Rényi divergences by using a cyclic refinement of the Jensen's inequality*, J. Bull. Malays. Math. Sci. Soc. 2017, https://doi.org/10.1007/s40840-017-0526-4.

[5] J. JAKŠETIĆ, Đ. PEČARIĆ AND J. PEČARIĆ, *Some properties of Zipf-Mandelbrot law and Hurwitz ζ-function*, Math. Inequal. Appl. 21 (2), 2018, 575–584.

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

[7] S. KULLBACK, *Information Theory and Statistics*, J. Wiley, New York, 1959.

[8] S. KULLBACK AND R. A.LEIBLER, *On information and sufficiency*, Annals Math. Statist. 12, 1951, 79–86.

[9] F. LIESE AND I. VAJDA, *Convex Statistical Distances (Teubner-Texte Zur Mathematik), vol. 95*, Teubner, Leipzig, Germany, 1987.

[10] R. MIKIĆ, Đ. PEČARIĆ AND J. PEČARIĆ, *Inequalities of the Jensen and Edmundson-Lah-Ribarič type for 3-convex functions with applications*, J. Math. Inequal. 12, 3, 2018, 677–692, dx.doi.org/10.7153/jmi-2018-12-52.

[11] Đ. PEČARIĆ, J. PEČARIĆ AND D. POKAZ, *Generalized Csiszár's $f-$divergence for Lipschitzian functions*, Math. Inequal. Appl., accepted.

[12] Đ. PEČARIĆ, J. PEČARIĆ AND M. RODIĆ, *About the sharpness of the Jensen inequality*, J. Inequal. Appl. (2018), 2018:337, https://doi.org/10.1186/s13660-018-1923-4.

[13] J. PEČARIĆ, I. PERIĆ AND M. RODIĆ LIPANOVIĆ, *Uniform treatment of Jensen type inequalities*, Math. Reports 16 (66), 2, 2014, 183–205.

[14] J. PEČARIĆ AND M. RODIĆ, *Uniform treatment of Jensen type inequalities II*, Math. Reports, 21 (71), 3, 2019.

[15] A. RENYI, *On measures of entropy and information*, Proc. Fourth Berkeley Symp. Math. Statist. Prob. Vol 1, University of California Press, Berkeley, 1961.

[16] I. VAJDA, *Theory of Statistical Inference and Information*, Kluwer, Dordrecht, 1989.

(Received November 4, 2018)

Đilda Pečarić
Catholic University of Croatia
Ilica 242, 10000 Zagreb, Croatia
e-mail: gildapeca@gmail.com

Josip Pečarić
Faculty of Textile Technology
University of Zagreb
Prilaz baruna Filipovića 28a, 10000 Zagreb, Croatia
RUDN University
Miklukho-Maklaya str. 6, 117198 Moscow, Russia
e-mail: pecaric@element.hr

Mirna Rodić
Faculty of Textile Technology
University of Zagreb
Prilaz baruna Filipovića 28a, 10000 Zagreb, Croatia
e-mail: mrodic@ttf.hr