Mathematical
Inequalities
& Applications
Volume 26, Number 2 (2023), 351–368 doi:10.7153/mia-2023-26-23

SHERMAN'S FUNCTIONAL, ITS PROPERTIES WITH APPLICATIONS FOR f–DIVERGENCE MEASURE

SLAVICA IVELIĆ BRADANOVIĆ*, ĐILDA PEČARIĆ AND JOSIP PEČARIĆ

(*Communicated by I. Perić*)

Abstract. In this paper we define Sherman's functional deduced from Sherman's inequality. We established lower and upper bounds for Sherman's functional and study its properties. As consequences of main results we obtained new bounds for Csiszár f-divergence functional and as special cases bounds for Shannon's entropy. As applications we use the Zipf-Mandelbrot law to introduce a new entropy and to derive some related results.

1. Introduction

Let I be an interval in $\mathbb{R}$. The well known Jensen inequality

$$f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \leqslant \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) \tag{1}$$

holds for every convex function $f : I \to \mathbb{R}$, for any vector $\mathbf{x} = (x_1,\dots,x_n) \in I^n$ and nonnegative n-tuple $\mathbf{p} = (p_1,\dots,p_n)$ with $\sum_{i=1}^{n} p_i = P_n > 0$.

If it is satified

$$p_1 > 0, \quad p_2,\dots,p_n \leqslant 0, \quad P_n > 0 \quad \text{and} \quad \frac{1}{P_n}\sum_{i=1}^{n} p_i x_i \in I,$$

then the reverse of Jensen's inequality

$$f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \geqslant \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i)$$

holds (see for example [32]).

Investigating the method of interpolating inequalities, J. Pečarić using Jensen's inequality and its reverse proved the following result ([22, p. 717]). This result was nor published in any journal and due that we give it with a proof.

Mathematics subject classification (2020): 94A17, 26D15, 15B51.
Keywords and phrases: Jensen's functional, Sherman's inequality, Csiszár f-divergence, Shannon's entropy, Zipf-Mandelbrot law.

* Corresponding author.

Paper MIA-26-23

THEOREM 1. *Let* $f : I \to \mathbb{R}$ *be a convex function,* $\mathbf{x} = (x_1, \ldots, x_n) \in I^n$, $\mathbf{p} = (p_1, \ldots, p_n)$ *and* $\mathbf{q} = (q_1, \ldots, q_n)$ *be nonnegative* n*-tuples such* $\mathbf{p} \geqslant \mathbf{q}$, *i.e.* $p_i \geqslant q_i$, $i = 1, \ldots, n$, *and* $P_n = \sum_{i=1}^n p_i > 0$, $Q_n = \sum_{i=1}^n q_i > 0$. *Then*

$$\sum_{i=1}^{n} q_i f(x_i) - Q_n f\left(\frac{1}{Q_n}\sum_{i=1}^{n} q_i x_i\right) \leqslant \sum_{i=1}^{n} p_i f(x_i) - P_n f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right). \tag{2}$$

Proof. Applying Jensen's inequality we have

$$\begin{aligned} &\sum_{i=1}^{n}(p_i - q_i) f\left(\frac{1}{\sum\limits_{i=1}^{n}(p_i - q_i)}\sum_{i=1}^{n}(p_i - q_i)x_i\right) \\ &\leqslant \sum_{i=1}^{n}(p_i - q_i)\frac{1}{\sum\limits_{i=1}^{n}(p_i - q_i)}\sum_{i=1}^{n}(p_i - q_i) f(x_i) = \sum_{i=1}^{n}(p_i - q_i) f(x_i). \end{aligned} \tag{3}$$

On the other side, applying the reverse of Jensen's inequality we get

$$\begin{aligned} &\sum_{i=1}^{n}(p_i - q_i) f\left(\frac{1}{\sum\limits_{i=1}^{n}(p_i - q_i)}\sum_{i=1}^{n}(p_i - q_i)x_i\right) \\ &= (P_n - Q_n) f\left(\frac{P_n\left(\frac{1}{P_n}\sum\limits_{i=1}^{n} p_i x_i\right) - Q_n\left(\frac{1}{Q_n}\sum\limits_{i=1}^{n} q_i x_i\right)}{P_n - Q_n}\right) \\ &\geqslant (P_n - Q_n)\left(\frac{P_n}{P_n - Q_n} f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) - \frac{Q_n}{P_n - Q_n} f\left(\frac{1}{Q_n}\sum_{i=1}^{n} q_i x_i\right)\right) \\ &= P_n f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) - Q_n f\left(\frac{1}{Q_n}\sum_{i=1}^{n} q_i x_i\right). \end{aligned} \tag{4}$$

Now, combining (3) and (4), we obtain (2). □

As an easy consequence of the previous theorem the following lemma holds.

LEMMA 1. *Let* $f : I \to \mathbb{R}$ *be a convex function,* $\mathbf{x} = (x_1, \ldots, x_n) \in I^n$ *and* $\mathbf{p} = (p_1, \ldots, p_n)$ *be nonnegative* n*-tuple such that* $P_n = \sum_{i=1}^n p_i > 0$. *Then*

$$\begin{aligned} \min_{1 \leqslant i \leqslant n}\{p_i\}\left[\sum_{i=1}^{n} f(x_i) - n f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\right] &\leqslant \sum_{i=1}^{n} p_i f(x_i) - P_n f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \\ &\leqslant \max_{1 \leqslant i \leqslant n}\{p_i\}\left[\sum_{i=1}^{n} f(x_i) - n f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\right]. \end{aligned} \tag{5}$$

Few years later, Dragomir et el. [11], obtained the analogous result to the previous lemma but as a consequence of a quite different approach. They considered Jensen's functional $\mathscr{J}_n(f,\mathbf{x},\mathbf{p})$, deduced from Jensen's inequality (1), as a difference between the right and left side, define as follows

$$\mathscr{J}_n(f,\mathbf{x},\mathbf{p}) = \sum_{i=1}^{n} p_i f(x_i) - P_n f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \geqslant 0. \tag{6}$$

From (6), for $\mathbf{p} = \mathbf{1} = (1,\ldots,1)$, we get

$$\mathscr{J}_n(f,\mathbf{x},\mathbf{1}) = \sum_{i=1}^{n} f(x_i) - nf\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \geqslant 0. \tag{7}$$

Dragomir continued the investigation of the properties of normalized Jensen's functional, i.e. functional (6) satisfying $P_n = \sum_{i=1}^{n} p_i = 1$, and obtained the following lower and upper bound for the normalized functional (see [9, Theorem 1]).

THEOREM 2. *Let* $f : I \to \mathbb{R}$ *be a convex function,* $\mathbf{x} = (x_1,\ldots,x_n) \in I^n$, $\mathbf{p} = (p_1,\ldots,p_n)$ *be a nonnegative* n*-tuple and* $\mathbf{q} = (q_1,\ldots,q_n)$ *be a positive* n*-tuple with* $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$. *Then*

$$(0 \leqslant) \min_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\} \mathscr{J}_n(f,\mathbf{x},\mathbf{q}) \leqslant \mathscr{J}_n(f,\mathbf{x},\mathbf{p}) \leqslant \max_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\} \mathscr{J}_n(f,\mathbf{x},\mathbf{q}). \tag{8}$$

We also recommend to the reader the monograph [20], in which results related to the concept of Jensen's functional are collected.

2. Preliminaries

For two vectors $\mathbf{x} = (x_1,\ldots,x_m)$, $\mathbf{y} = (y_1,\ldots,y_m) \in I^m$, where $x_{[i]}, y_{[i]}$ denote their increasing order, we say that $\mathbf{x}$ majorizes $\mathbf{y}$ and write $\mathbf{y} \prec \mathbf{x}$ if

$$\sum_{i=1}^{k} y_{[i]} \leqslant \sum_{i=1}^{k} x_{[i]} \quad k = 1,\ldots,m, \quad \text{and} \quad \sum_{i=1}^{m} y_i = \sum_{i=1}^{m} x_i. \tag{9}$$

Let $\mathscr{M}_{nm}(\mathbb{R}_+)$ denotes the set of all $n \times m$ matrices with nonnegative real entries. A matrix $A \in \mathscr{M}_{nm}(\mathbb{R}_+)$ is doubly stochastic matrix if the sum of the entries in each column and row is equal to 1. A matrix $A \in \mathscr{M}_{nm}(\mathbb{R}_+)$ is column (resp. row) stochastic matrix if the sum of the entries in each column (resp. row) is equal to 1. With $\mathscr{C}_{nm}(\mathbb{R}_+)$ we denote the set of all $n \times m$ column stochastic matrices with nonnegative real entries. For column stochastic matrices with positive real entries we use notation $\mathscr{C}_{nm}((0,\infty))$.

It is well known that

$$\mathbf{y} \prec \mathbf{x} \quad \text{iff} \quad \mathbf{y} = \mathbf{x}\mathbf{A}$$

for some doubly stochastic matrix $\mathbf{A} \in \mathscr{M}_{mm}(\mathbb{R}_+)$.

Moreover, for every convex function $f: I \to \mathbb{R}$, the relation $\mathbf{y} \prec \mathbf{x}$ implies the well known Majorization inequality

$$\sum_{i=1}^{m} f(y_i) \leqslant \sum_{i=1}^{m} f(x_i). \tag{10}$$

Sherman [34] considered the weighted concept of majorization

$$(\mathbf{y}, \mathbf{b}) \prec (\mathbf{x}, \mathbf{a}),$$

between two vectors $\mathbf{x} = (x_1, \ldots, x_n) \in I^n$ and $\mathbf{y} = (y_1, \ldots, y_m) \in I^m$ with nonnegative weights $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_m)$ under the assumption of existence of column stochastic matrix $\mathbf{S} = (a_{ij}) \in \mathscr{C}_{nm}(\mathbb{R}_+)$ such that

$$\mathbf{y} = \mathbf{x}\mathbf{S} \quad \text{and} \quad \mathbf{a} = \mathbf{b}\mathbf{S}^T, \tag{11}$$

where $\mathbf{S}^T$ is a transpose matrix of $\mathbf{S}$. Under conditions (11), for every convex function $f: I \to \mathbb{R}$, the following inequality

$$\sum_{j=1}^{m} b_j f(y_j) \leqslant \sum_{i=1}^{n} a_i f(x_i) \tag{12}$$

holds. This result is known as Sherman's theorem and (12) as Sherman's inequality. It represents a generalization of Majorization inequality (10) as well as Jensen's inequality (1). Choosing $m = n$ and $\mathbf{b} = \mathbf{1} = (1, \ldots, 1)$, the condition $(\mathbf{y}, \mathbf{b}) \prec (\mathbf{x}, \mathbf{a})$ implies $\mathbf{y} = \mathbf{x}\mathbf{S}$ with some doubly stochastic matrix $\mathbf{S}$ and $\mathbf{a} = \mathbf{b}\mathbf{S}^T = \mathbf{1} = (1, \ldots, 1)$, and as direct consequence of Sherman's inequality we get Majorization inequality. Specially, choosing $m = 1$ and setting $b_1 = 1$, $s_{i1} = \frac{p_i}{P_n}$, Sherman's inequality (12) reduces to Jensen's inequality (1). Results obtained for Sherman's inequality generalize certain results that are valid for Majorization and Jensen's inequality. Some resent generalizations of Sherman's inequality can be found in [1]–[4], [6], [12]–[19], [25]–[31].

Our paper is organized as follows. In the third section, we define Sherman's functional. We study its properties and establish lower and upper bounds for it. In the fourth section, we obtain new lower and upper bounds for Csiszár f-divergence functionals and Shannon's entropy. The last section is dedicated to the Zipf-Mandelbrot law. In combination with results from the previous section, we introduce a new entropy and derive some related results.

3. Sherman's functional and its properties

Motivated by Sherman's inequality (12), i.e.

$$0 \leqslant \sum_{i=1}^{n} a_i f(x_i) - \sum_{j=1}^{m} b_j f(y_j) \tag{13}$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} b_j s_{ij} \right) f(x_i) - \sum_{j=1}^{m} b_j f\left(\sum_{i=1}^{n} x_i s_{ij} \right), \tag{14}$$

we define Sherman's functional as follows.

Let $\mathscr{F}(I)$ denotes the set of all convex functions defined on I. Under the assumptions of Sherman's theorem, we define Sherman's functional $\mathscr{S} : \mathscr{F}(I) \times I^n \times [0,\infty)^m \times \mathscr{C}_{nm}(\mathbb{R}_+) \to \mathbb{R}$ by

$$\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S}) = \sum_{i=1}^{n}\sum_{j=1}^{m} b_j s_{ij} f(x_i) - \sum_{j=1}^{m} b_j f\left(\sum_{i=1}^{n} x_i s_{ij}\right) \tag{15}$$
$$= \sum_{j=1}^{m} b_j \left(\sum_{i=1}^{n} s_{ij} f(x_i) - f\left(\sum_{i=1}^{n} x_i s_{ij}\right)\right).$$

THEOREM 3. *Sherman's functional $\mathscr{S}$ satisfies the following properties:*

a) $\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S})$ *is positive, i.e.*

$$\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S}) \geqslant 0; \tag{16}$$

b) $\mathscr{S}(f,\mathbf{x},\mathbf{b},\cdot)$ *is concave on the convex set* $\mathscr{C}_{nm}(\mathbb{R}_+)$, *i.e. for all* $\lambda \in [0,1]$

$$\mathscr{S}(f,\mathbf{x},\mathbf{b},\lambda\mathbf{S}+(1-\lambda)\mathbf{T}) \geqslant \lambda\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S}) + (1-\lambda)\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{T}); \tag{17}$$

c) $\mathscr{S}(\mathbf{x},\cdot,\mathbf{S})$ *is linear mapping, i.e.*

$$\mathscr{S}(f,\mathbf{x},\mathbf{b}+\mathbf{d},\mathbf{S}) = \mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S}) + \mathscr{S}(f,\mathbf{x},\mathbf{d},\mathbf{S}), \tag{18}$$
$$\mathscr{S}(f,\mathbf{x},\lambda\mathbf{b},\mathbf{S}) = \lambda\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S});$$

d) $\mathscr{S}(f,\mathbf{x},\cdot,\mathbf{S})$ *is increasing on* $[0,\infty)^m$, *i.e. if* $\mathbf{b},\mathbf{c} \in [0,\infty)^m$ *such that* $\mathbf{b} \geqslant \mathbf{d}$, *i.e.* $b_j \geqslant d_j, j = 1,\ldots,m$, *then*

$$\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S}) \geqslant \mathscr{S}(f,\mathbf{x},\mathbf{d},\mathbf{S}) \geqslant 0. \tag{19}$$

Proof. a) Positivity of functional $\mathscr{S}$ follows from Sherman's inequality (12).

b) We have

$$\mathscr{S}(f,\mathbf{x},\mathbf{b},\lambda\mathbf{S}+(1-\lambda)\mathbf{T}) \tag{20}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} b_j(\lambda s_{ij}+(1-\lambda)t_{ij}) f(x_i) - \sum_{j=1}^{m} b_j f\left(\sum_{i=1}^{n}(\lambda s_{ij}+(1-\lambda)t_{ij})x_i\right)$$
$$= \lambda\sum_{i=1}^{n}\sum_{j=1}^{m} b_j s_{ij} f(x_i) + (1-\lambda)\sum_{i=1}^{n}\sum_{j=1}^{m} b_j t_{ij} f(x_i) - \sum_{j=1}^{m} b_j f\left(\sum_{i=1}^{n}(\lambda s_{ij}+(1-\lambda)t_{ij})x_i\right).$$

Further, by convexity of f we have

$$\begin{aligned}
&\sum_{j=1}^{m} b_j f\left(\sum_{i=1}^{n}(\lambda s_{ij}+(1-\lambda)t_{ij})x_i\right) \\
&\leqslant \sum_{j=1}^{m} b_j\left[\lambda f\left(\sum_{i=1}^{n} s_{ij}x_i\right)+(1-\lambda)f\left(\sum_{i=1}^{n} t_{ij}x_i\right)\right] \\
&=\lambda\sum_{j=1}^{m} b_j f\left(\sum_{i=1}^{n} s_{ij}x_i\right)+(1-\lambda)\sum_{j=1}^{m} b_j f\left(\sum_{i=1}^{n} t_{ij}x_i\right).
\end{aligned} \tag{21}$$

Combining (20) and (21), we prove the concavity property b).

c) It is obvious.

d) Case $\mathbf{b}=\mathbf{c}$ is trivial. Let $\mathbf{b}>\mathbf{c}$. Since

$$\mathbf{b}=(\mathbf{b}-\mathbf{c})+\mathbf{c},$$

using the property linearity (18) and positivity (16), we have

$$\begin{aligned}
\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S}) &= \mathscr{S}(f,\mathbf{x},(\mathbf{b}-\mathbf{c})+\mathbf{c},\mathbf{S}) \\
&= \mathscr{S}(f,\mathbf{x},\mathbf{b}-\mathbf{c},\mathbf{S})+\mathscr{S}(f,\mathbf{x},\mathbf{c},\mathbf{S}) \\
&\geqslant \mathscr{S}(f,\mathbf{x},\mathbf{c},\mathbf{S})\geqslant 0,
\end{aligned}$$

what we need to prove. This ends the proof. □

In the following theorems we establish lower and upper bounds for Sherman's functional.

THEOREM 4. *For* $\mathbf{b}=(b_1,\ldots,b_m)\in[0,\infty)^m$ *and* $\mathbf{S}=(s_{ij})\in\mathscr{C}_{nm}(\mathbb{R}_+)$, *we define*

$$\begin{aligned}
\tilde{m} &= \sum_{j=1}^{m} b_j \min_{1\leqslant i\leqslant n}\{s_{ij}\}, \\
\tilde{M} &= \sum_{j=1}^{m} b_j \max_{1\leqslant i\leqslant n}\{s_{ij}\}.
\end{aligned}$$

Then

$$0\leqslant \tilde{m}\cdot\mathscr{J}_n(f,\mathbf{x},\mathbf{1})\leqslant\mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S})\leqslant\tilde{M}\cdot\mathscr{J}_n(f,\mathbf{x},\mathbf{1}). \tag{22}$$

Proof. The first inequality in (22) is consequence of nonnegativity of all entries of $\mathbf{b}$, $\mathbf{S}$ and Jensen's functional $\mathscr{J}_n(f,\mathbf{x},\mathbf{1})$.

Again this is a consequence of Theorem 1. We present an alternative approach (analogous to [9]).

We prove

$$\min_{1\leqslant i\leqslant n}\{s_{ij}\}\cdot\mathscr{J}_n(f,\mathbf{x},\mathbf{1})\leqslant\sum_{i=1}^{n} s_{ij}f(x_i)-f\left(\sum_{i=1}^{n} x_i s_{ij}\right)\leqslant\max_{1\leqslant i\leqslant n}\{s_{ij}\}\cdot\mathscr{J}_n(f,\mathbf{x},\mathbf{1}), \tag{23}$$

for all $j = 1,\dots,m$.

Let us denote $\tilde{m}_j = \min_{1\leqslant i\leqslant n}\{s_{ij}\}$ and $\tilde{M}_j = \max_{1\leqslant i\leqslant n}\{s_{ij}\}$, $j \in \{1,\dots,m\}$.

Then (23) can be written as

$$\tilde{m}_j \cdot \mathscr{J}_n(f,\mathbf{x},\mathbf{1}) \leqslant \sum_{i=1}^{n} s_{ij} f(x_i) - f\left(\sum_{i=1}^{n} x_i s_{ij}\right) \leqslant \tilde{M}_j \cdot \mathscr{J}_n(f,\mathbf{x},\mathbf{1}). \tag{24}$$

Since $n\cdot\tilde{m}_j + \sum_{i=1}^{n}(s_{ij} - \tilde{m}_j) = 1$, applying Jensen's inequality we have

$$\begin{aligned}
f\left(\sum_{i=1}^{n} x_i s_{ij}\right) &= f\left(n\cdot\tilde{m}_j\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + \sum_{i=1}^{n}(s_{ij}-\tilde{m}_j)x_i\right) \\
&\leqslant n\cdot\tilde{m}_j f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + \sum_{i=1}^{n}(s_{ij}-\tilde{m}_j) f(x_i) \\
&= \sum_{i=1}^{n} s_{ij} f(x_i) - \tilde{m}_j\left(\sum_{i=1}^{n} f(x_i) - nf\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\right) \\
&= \sum_{i=1}^{n} s_{ij} f(x_i) - \tilde{m}_j \cdot \mathscr{J}_n(f,\mathbf{x},\mathbf{1})
\end{aligned}$$

which gives the left hand side of (24).

Further, since $\tilde{M}_j > 0$ and $\frac{1}{\tilde{M}_j}\sum_{i=1}^{n}\left(\frac{\tilde{M}_j}{n} - \frac{s_{ij}}{n}\right) + \frac{1}{n\tilde{M}_j} = 1$, applying Jensen's inequality we have

$$\begin{aligned}
f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) &= f\left(\frac{1}{\tilde{M}_j}\sum_{i=1}^{n}\left(\frac{\tilde{M}_j}{n} - \frac{s_{ij}}{n}\right)x_i + \frac{1}{n\tilde{M}_j}\left(\sum_{i=1}^{n} s_{ij}x_i\right)\right) \\
&\leqslant \frac{1}{\tilde{M}_j}\sum_{i=1}^{n}\left(\frac{\tilde{M}_j}{n} - \frac{s_{ij}}{n}\right) f(x_i) + \frac{1}{n\tilde{M}_j} f\left(\sum_{i=1}^{n} s_{ij}x_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} f(x_i) - \frac{1}{n\tilde{M}_j}\left(\sum_{i=1}^{n} s_{ij} f(x_i) - f\left(\sum_{i=1}^{n} s_{ij}x_i\right)\right),
\end{aligned}$$

which gives the right hand side of (24).

Therefore, the inequality (23) holds. Now, multiplying with b_j and summing over $j = 1,\dots,m$, we get (22) what we need to prove. □

THEOREM 5. *For* $\mathbf{b} = (b_1,\dots,b_m) \in [0,\infty)^m$, $\mathbf{S} = (s_{ij}) \in \mathscr{C}_{nm}(\mathbb{R}_+)$ *and* $\mathbf{T} = (t_{ij}) \in \mathscr{C}_{nm}((0,\infty))$, *we define*

$$\begin{aligned}
\bar{\mathbf{b}}_{\min} &= \left(b_1 \min_{1\leqslant i\leqslant n}\left\{\frac{s_{i1}}{t_{i1}}\right\},\dots,b_m \min_{1\leqslant i\leqslant n}\left\{\frac{s_{im}}{t_{im}}\right\}\right), \\
\bar{\mathbf{b}}_{\max} &= \left(b_1 \max_{1\leqslant i\leqslant n}\left\{\frac{s_{i1}}{t_{i1}}\right\},\dots,b_m \max_{1\leqslant i\leqslant n}\left\{\frac{s_{im}}{t_{im}}\right\}\right).
\end{aligned}$$

Then

$$(0 \leqslant)\ \mathscr{S}(f,\mathbf{x},\overline{\mathbf{b}}_{\min},\mathbf{T}) \leqslant \mathscr{S}(f,\mathbf{x},\mathbf{b},\mathbf{S}) \leqslant \mathscr{S}(f,\mathbf{x},\overline{\mathbf{b}}_{\max},\mathbf{T}). \tag{25}$$

Proof. According to the notation, we need to prove

$$\begin{aligned}
&\sum_{j=1}^{m} b_j \min_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\}\left(\sum_{i=1}^{n} t_{ij} f(x_i) - f\left(\sum_{i=1}^{n} x_i t_{ij}\right)\right)\\
&\leqslant \sum_{j=1}^{m} b_j \left(\sum_{i=1}^{n} s_{ij} f(x_i) - f\left(\sum_{i=1}^{n} x_i s_{ij}\right)\right)\\
&\leqslant \sum_{j=1}^{m} b_j \max_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\}\left(\sum_{i=1}^{n} t_{ij} f(x_i) - f\left(\sum_{i=1}^{n} x_i t_{ij}\right)\right)
\end{aligned}$$

Let us consider n-tuples $\mathbf{p}_j$ and $\mathbf{q}_{\min,j}$ such that

$$\begin{aligned}
\mathbf{p}_j &= \left(t_{1j}\frac{s_{1j}}{t_{1j}},\ldots,t_{nj}\frac{s_{nj}}{t_{nj}}\right) = (s_{1j},\ldots,s_{nj}),\\
\mathbf{q}_{\min,j} &= \left(t_{1j}\min_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\},\ldots,t_{nj}\min_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\}\right).
\end{aligned}$$

It is obviously $\mathbf{q}_{\min,j} \leqslant \mathbf{p}_j$. Applying Theorem 1 to n-tuples $\mathbf{q}_{\min,j}$ and $\mathbf{p}_j$, we get

$$\begin{aligned}
&\sum_{i=1}^{n} s_{ij} f(x_i) - \sum_{i=1}^{n} s_{ij} f\left(\frac{1}{\sum_{i=1}^{n} s_{ij}}\sum_{i=1}^{n} s_{ij} x_i\right)\\
&\geqslant \min_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\}\left[\sum_{i=1}^{n} t_{ij} f(x_i) - f\left(\sum_{i=1}^{n} t_{ij} x_i\right)\right],
\end{aligned}$$

i.e. we get

$$\sum_{i=1}^{n} s_{ij} f(x_i) - f\left(\sum_{i=1}^{n} s_{ij} x_i\right) \geqslant \min_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\}\left[\sum_{i=1}^{n} t_{ij} f(x_i) - f\left(\sum_{i=1}^{n} t_{ij} x_i\right)\right].$$

Exchanging the roles of *min* and *max*, we get the inequality

$$\sum_{i=1}^{n} s_{ij} f(x_i) - f\left(\sum_{i=1}^{n} s_{ij} x_i\right) \leqslant \max_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\}\left[\sum_{i=1}^{n} t_{ij} f(x_i) - f\left(\sum_{i=1}^{n} t_{ij} x_i\right)\right].$$

So, we have

$$\begin{aligned}
&\min_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\}\left[\sum_{i=1}^{n} t_{ij} f(x_i) - f\left(\sum_{i=1}^{n} t_{ij} x_i\right)\right] \\
&\leqslant \sum_{i=1}^{n} s_{ij} f(x_i) - f\left(\sum_{i=1}^{n} s_{ij} x_i\right) \\
&\leqslant \max_{1\leqslant i\leqslant n}\left\{\frac{s_{ij}}{t_{ij}}\right\}\left[\sum_{i=1}^{n} t_{ij} f(x_i) - f\left(\sum_{i=1}^{n} t_{ij} x_i\right)\right].
\end{aligned}$$

Now multiplying with b_j and summing over $j = 1,\ldots,m$, we get the required result. □

REMARK 1. Choosing $m = 1$ and setting $b_1 = 1$ and $s_{i1} = \frac{p_i}{P_n}$, for $i = 1,\ldots,n$, Sherman's functional (15) reduces to Jensen's functional that corresponds to (6). Therefore, applying Theorem 4, we obtain the bounds of Jensen's functional in the form (5), i.e.

$$0 \leqslant \min_{1\leqslant i\leqslant n}\{p_i\}\mathscr{J}_n(f,\mathbf{x},\mathbf{1}) \leqslant \mathscr{J}_n(f,\mathbf{x},\mathbf{p}) \leqslant \max_{1\leqslant i\leqslant n}\{p_i\}\mathscr{J}_n(f,\mathbf{x},\mathbf{1}).$$

Applying Theorem 3, the concavity property b), for $\lambda = \frac{1}{2}$, with $m = 1$, $b_1 = 1$, $s_{i1} = p_i$, $t_{i1} = q_i$, for $i = 1,\ldots,n$, $P_n = Q_n = 1$, it implies the properties of Jensen's functional superadditivity

$$\mathscr{J}_n(f,\mathbf{x},\mathbf{p}+\mathbf{q}) \geqslant \mathscr{J}_n(f,\mathbf{x},\mathbf{p}) + \mathscr{J}_n(f,\mathbf{x},\mathbf{q}), \tag{26}$$

and monotonicity

$$\mathscr{J}_n(f,\mathbf{x},\mathbf{p}) \geqslant \mathscr{J}_n(f,\mathbf{x},\mathbf{q}) \geqslant 0, \quad \text{for } \mathbf{p} \geqslant \mathbf{q}. \tag{27}$$

Further, applying Theorem 5, with $m = 1$, $b_1 = 1$ and $s_{i1} = p_i$, $t_{i1} = q_i$, for $i = 1,\ldots,n$, the inequality (25) reduces to the inequality (8).

4. New bounds for f-divergence functional

Csiszár [7] introduced the concept of f-divergence functional

$$D_f(\mathbf{q},\mathbf{p}) = \sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right) \tag{28}$$

for a convex function $f : [0,\infty) \to \mathbb{R}$ and nonnegative n-tuples $\mathbf{p} = (p_1,\ldots,p_n)$, $\mathbf{q} = (q_1,\ldots,q_n)$ with the convention

$$\begin{aligned}
&f(0) = \lim_{t\to 0+} f(t), \quad \frac{0}{0} = 0, \quad 0f\left(\frac{0}{0}\right) = 0 \\
&0f\left(\frac{c}{0}\right) = \lim_{\varepsilon\to 0+} \varepsilon f\left(\frac{c}{\varepsilon}\right) = c\lim_{t\to\infty}\frac{f(t)}{t}, \quad c > 0.
\end{aligned} \tag{29}$$

For a different choice of the function f, as a special case of (28), we may obtain various divergences as the Kullback-Leibler distance, α-order Rényi entropy, Bhattacharyya distance, Harmonic distance, $\varkappa^2$-divergence, Hellinger distance etc. which play an important role in Information theory and Statistics. Many papers are devoted to the notion of f-divergence and the subject of inequalities for different types of divergences (see for example [4], [10], [12], [13], [15], [19], [33]).

In [19], the authors introduced the weighted Csiszár f-divergence functional, with weights $r_1,\dots,r_n \geqslant 0$, defined by

$$D_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right). \tag{30}$$

REMARK 2. For $\mathbf{1} = (1,\dots,1) \in \mathbb{R}^n$ we have $D_f(\mathbf{q},\mathbf{p};\mathbf{1}) = D_f(\mathbf{q},\mathbf{p})$.

The classical inequality for f-divergence functional, known as the Csiszár-Körner inequality, is given in the next theorem (see [8]).

THEOREM 6. *Let $f:[0,\infty) \to \mathbb{R}$ be a convex function, $\mathbf{p} = (p_1,\dots,p_n) \in [0,\infty)^n$ and $\mathbf{q} = (q_1,\dots,q_n) \in [0,\infty)^n$ with $P_n = \sum_{i=1}^{n} p_i > 0$ and $Q_n = \sum_{i=1}^{n} q_i > 0$. Then*

$$0 \leqslant D_f(\mathbf{q},\mathbf{p}) - P_n f\left(\frac{Q_n}{P_n}\right). \tag{31}$$

If f is normalized, i.e. $f(1) = 0$ and $P_n = Q_n$, then

$$D_f(\mathbf{q},\mathbf{p}) \geqslant 0. \tag{32}$$

In particular, if $\mathbf{p}$ and $\mathbf{q}$ are two probability distribution, then the inequality (32) holds for every convex and normalized function f.

Shannon [33] introduced a statistical concept of entropy, the measure of information

$$H(\mathbf{p}) = \sum_{i=1}^{n} p_i \ln \frac{1}{p_i}, \tag{33}$$

where $\mathbf{p} = (p_1,\dots,p_n)$ is a probability distribution for some discrete random variable X. It quantifies the unevenness in the probability distribution $\mathbf{p}$.

Motivated by various communication and transmission problems, Belis and Guiaşu [5] introduced the concept of weighted Shannon's entropy with nonnegative weights r_i, $i = 1,\dots,n$, defined by

$$H(\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i \ln \frac{1}{p_i}, \tag{34}$$

(see also [35]).

REMARK 3. If we ignore weights r_i, $i = 1,\dots,n$, then (34) reduces to (33), i.e. $H(\mathbf{p};\mathbf{1}) = H(\mathbf{p})$ for $\mathbf{1} = (1,\dots,1) \in \mathbb{R}^n$.

The weighted entropy satisfies the estimate

$$0 \leqslant H(\mathbf{p};\mathbf{r}) \leqslant \sum_{i=1}^{n} r_i p_i \ln \frac{\sum_{i=1}^{n} r_i}{\sum_{i=1}^{n} r_i p_i}$$

(see [24]). In particular, the minimum $H(\mathbf{p};\mathbf{r}) = 0$ is reached for a constant random variable, i.e. when $p_i = 1$, for some i. The opposite extreme, the maximal $H(\mathbf{p};\mathbf{r})$ is reached for a uniform distribution, i.e. when $p_i = \frac{1}{n}$ for all $i = 1,\dots,n$. In that case we have

$$0 \leqslant H(\mathbf{p};\mathbf{r}) \leqslant \frac{1}{n}\sum_{i=1}^{n} r_i \ln n.$$

Specially, ignoring weights r_i, $i = 1,\dots,n$, i.e. setting $\mathbf{r} = \mathbf{1} = (1,\dots,1)$, the previous inequality reduces to

$$0 \leqslant H(\mathbf{p}) \leqslant \ln n.$$

In the following results we use the convention (29) and notation $\langle \cdot,\cdot \rangle$ for the standard inner product.

As an easy consequence of Theorem 5 we get the next result.

COROLLARY 1. *Let* $f : [0,\infty) \to \mathbb{R}$ *be a convex function,* $\mathbf{p} = (p_1,\dots,p_n) \in [0,\infty)^n$, $\mathbf{q} = (q_1,\dots,q_n) \in [0,\infty)^n$ *and* $\mathbf{R} = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$. *Let us define*

$$\langle \mathbf{p},\mathbf{r}_j \rangle = \sum_{i=1}^{n} p_i r_{ij}, \quad \langle \mathbf{q},\mathbf{r}_j \rangle = \sum_{i=1}^{n} q_i r_{ij}, \quad j = 1,\dots,m, \tag{35}$$

$$R_f(\mathbf{q},\mathbf{p};\mathbf{R}) = \sum_{j=1}^{m}\left[\sum_{i=1}^{n} q_i r_{ij} f\left(\frac{q_i}{p_i}\right) - \langle \mathbf{q},\mathbf{r}_j \rangle f\left(\frac{1}{\langle \mathbf{q},\mathbf{r}_j \rangle}\sum_{i=1}^{n}\frac{q_i^2 r_{ij}}{p_i}\right)\right].$$

Then

$$\begin{aligned} 0 &\leqslant \min_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\} R_f(\mathbf{q},\mathbf{p};\mathbf{R}) \\ &\leqslant \sum_{j=1}^{m}\left(\sum_{i=1}^{n} p_i r_{ij} f\left(\frac{q_i}{p_i}\right) - \langle \mathbf{p},\mathbf{r}_j \rangle f\left(\frac{1}{\langle \mathbf{p},\mathbf{r}_j \rangle}\sum_{i=1}^{n} q_i r_{ij}\right)\right) \\ &\leqslant \max_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\} R_f(\mathbf{q},\mathbf{p};\mathbf{R}) \end{aligned} \tag{36}$$

Proof. Applying (25) to the vector $\mathbf{x} = (x_1,\dots,x_n)$ with $x_i = \frac{q_i}{p_i}$, $i = 1,\dots,n$, weight $\mathbf{b} = (b_1,\dots,b_m)$ with $b_j = \langle \mathbf{p},\mathbf{r}_j \rangle, j = 1,\dots,m$, column stochastic matrices $\mathbf{S} = (s_{ij})$ with $s_{ij} = \frac{p_i r_{ij}}{\langle \mathbf{p},\mathbf{r}_j \rangle}$, $i = 1,\dots,n$, $j = 1,\dots,m$ and $\mathbf{T} = (t_{ij})$ with $t_{ij} = \frac{q_i r_{ij}}{\langle \mathbf{q},\mathbf{r}_j \rangle}$,

$i = 1, \dots, n,\ j = 1, \dots, m$, we get

$$\begin{aligned}
0 &\leqslant \sum_{j=1}^{m} \langle \mathbf{p}, \mathbf{r}_j \rangle \min_{1\leqslant i\leqslant n} \left\{ \frac{\frac{p_i r_{ij}}{\langle \mathbf{p}, \mathbf{r}_j \rangle}}{\frac{q_i r_{ij}}{\langle \mathbf{q}, \mathbf{r}_j \rangle}} \right\} \left(\sum_{i=1}^{n} \frac{q_i r_{ij}}{\langle \mathbf{q}, \mathbf{r}_j \rangle} f\left(\frac{q_i}{p_i}\right) - f\left(\sum_{i=1}^{n} \frac{q_i}{p_i} \frac{q_i r_{ij}}{\langle \mathbf{q}, \mathbf{r}_j \rangle} \right) \right) \\
&\leqslant \sum_{j=1}^{m} \langle \mathbf{p}, \mathbf{r}_j \rangle \left(\sum_{i=1}^{n} \frac{p_i r_{ij}}{\langle \mathbf{p}, \mathbf{r}_j \rangle} f\left(\frac{q_i}{p_i}\right) - f\left(\sum_{i=1}^{n} \frac{q_i}{p_i} \frac{p_i r_{ij}}{\langle \mathbf{p}, \mathbf{r}_j \rangle} \right) \right) \\
&\leqslant \sum_{j=1}^{m} \langle \mathbf{p}, \mathbf{r}_j \rangle \max_{1\leqslant i\leqslant n} \left\{ \frac{\frac{p_i r_{ij}}{\langle \mathbf{p}, \mathbf{r}_j \rangle}}{\frac{q_i r_{ij}}{\langle \mathbf{q}, \mathbf{r}_j \rangle}} \right\} \left(\sum_{i=1}^{n} \frac{q_i r_{ij}}{\langle \mathbf{q}, \mathbf{r}_j \rangle} f\left(\frac{q_i}{p_i}\right) - f\left(\sum_{i=1}^{n} \frac{q_i}{p_i} \frac{q_i r_{ij}}{\langle \mathbf{q}, \mathbf{r}_j \rangle} \right) \right)
\end{aligned}$$

which is equivalent to (36). □

Specially, for $m = 1$, the previous result reduces to the next corollary.

COROLLARY 2. *Let* $f : [0,\infty) \to \mathbb{R}$ *be a convex function,* $\mathbf{p} = (p_1, \dots, p_n) \in [0,\infty)^n$, $\mathbf{q} = (q_1, \dots, q_n) \in [0,\infty)^n$ *and* $\mathbf{r} = (r_1, \dots, r_n) \in [0,\infty)^n$. *Let us define*

$$\begin{aligned}
\langle \mathbf{p}, \mathbf{r} \rangle &= \sum_{i=1}^{n} p_i r_i, \quad \langle \mathbf{q}, \mathbf{r} \rangle = \sum_{i=1}^{n} q_i r_i, \\
R_f(\mathbf{q}, \mathbf{p}; \mathbf{r}) &= \sum_{i=1}^{n} q_i r_i f\left(\frac{q_i}{p_i}\right) - \langle \mathbf{q}, \mathbf{r} \rangle f\left(\frac{1}{\langle \mathbf{q}, \mathbf{r} \rangle} \sum_{i=1}^{n} \frac{q_i^2 r_i}{p_i} \right).
\end{aligned}$$

Then

$$\begin{aligned}
0 &\leqslant \min_{1\leqslant i\leqslant n} \left\{ \frac{p_i}{q_i} \right\} R_f(\mathbf{q}, \mathbf{p}; \mathbf{r}) \leqslant D_f(\mathbf{q}, \mathbf{p}; \mathbf{r}) - \langle \mathbf{p}, \mathbf{r} \rangle f\left(\frac{1}{\langle \mathbf{p}, \mathbf{r} \rangle} \sum_{i=1}^{n} q_i r_i \right) \\
&\leqslant \max_{1\leqslant i\leqslant n} \left\{ \frac{p_i}{q_i} \right\} R_f(\mathbf{q}, \mathbf{p}; \mathbf{r}). \qquad (37)
\end{aligned}$$

If in addition $\mathbf{r} = \mathbf{1} = (1, \dots, 1)$, *with notations* $\langle \mathbf{p}, \mathbf{r} \rangle = \sum_{i=1}^{n} p_i = P_n$, $\langle \mathbf{p}, \mathbf{r} \rangle = \sum_{i=1}^{n} q_i = Q_n$, *then*

$$\begin{aligned}
0 &\leqslant \min_{1\leqslant i\leqslant n} \left\{ \frac{p_i}{q_i} \right\} \left(\sum_{i=1}^{n} q_i f\left(\frac{q_i}{p_i}\right) - Q_n f\left(\frac{1}{Q_n} \sum_{i=1}^{n} \frac{q_i^2}{p_i} \right) \right) \qquad (38) \\
&\leqslant D_f(\mathbf{q}, \mathbf{p}) - P_n f\left(\frac{Q_n}{P_n}\right) \\
&\leqslant \max_{1\leqslant i\leqslant n} \left\{ \frac{p_i}{q_i} \right\} \left(\sum_{i=1}^{n} q_i f\left(\frac{q_i}{p_i}\right) - Q_n f\left(\frac{1}{Q_n} \sum_{i=1}^{n} \frac{q_i^2}{p_i} \right) \right).
\end{aligned}$$

REMARK 4. Note that the results from the previous corollary generalize and refine the Csizar-Korner inequality (31). Specially, applying (38) to normalized function f, i.e. $f(1)=0$, with $P_n=Q_n$, we get the lower and upper bound of $D_f(\mathbf{q},\mathbf{p})$ in the form

$$\begin{aligned} 0 &\leqslant \min_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\}\left(\sum_{i=1}^{n} q_i f\left(\frac{q_i}{p_i}\right)-Q_n f\left(\frac{1}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i}\right)\right)\\ &\leqslant D_f(\mathbf{q},\mathbf{p})\\ &\leqslant \max_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\}\left(\sum_{i=1}^{n} q_i f\left(\frac{q_i}{p_i}\right)-Q_n f\left(\frac{1}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i}\right)\right). \end{aligned}$$

In the proofs of the following two corollaries we use the next conclusions.

REMARK 5. Choosing the convex function $f(t)=-\ln t$, we have

$$\begin{aligned} D_f(\mathbf{q},\mathbf{p}) &= -\sum_{i=1}^{n} p_i \ln\left(\frac{q_i}{p_i}\right)=\sum_{i=1}^{n} p_i\ln\frac{1}{q_i}-H(\mathbf{p}),\\ D_f(\mathbf{q},\mathbf{p};\mathbf{r}) &= -\sum_{i=1}^{n} r_i p_i \ln\left(\frac{q_i}{p_i}\right)=\sum_{i=1}^{n} r_i p_i\ln\frac{1}{q_i}-H(\mathbf{p};\mathbf{r}). \end{aligned}$$

Setting $\mathbf{q}=\mathbf{1}=(1,\ldots,1)$, we get

$$D_f(\mathbf{1},\mathbf{p})=-H(\mathbf{p}) \quad\text{and}\quad D_f(\mathbf{1},\mathbf{p};\mathbf{r})=-H(\mathbf{p};\mathbf{r}).$$

We estimate new bounds for Shannon's entropy.

COROLLARY 3. *Let* $\mathbf{p}=(p_1,\ldots,p_n)$ *be a probability distribution and* $\mathbf{q}=(q_1,\ldots,q_n)\in[0,\infty)^n$, $\mathbf{r}=(r_1,\ldots,r_n)\in[0,\infty)^n$. *Then*

$$\begin{aligned} &\langle\mathbf{p},\mathbf{r}\rangle\ln\left(\frac{1}{\langle\mathbf{p},\mathbf{r}\rangle}\sum_{i=1}^{n} q_i r_i\right)+\sum_{i=1}^{n} r_i p_i\ln\frac{1}{q_i}-\max_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\}R_{-\ln}(\mathbf{q},\mathbf{p};\mathbf{r})\\ &\leqslant H(\mathbf{p};\mathbf{r}) \qquad\qquad (39)\\ &\leqslant \langle\mathbf{p},\mathbf{r}\rangle\ln\left(\frac{1}{\langle\mathbf{p},\mathbf{r}\rangle}\sum_{i=1}^{n} q_i r_i\right)+\sum_{i=1}^{n} r_i p_i\ln\frac{1}{q_i}-\min_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\}R_{-\ln}(\mathbf{q},\mathbf{p};\mathbf{r})\\ &\leqslant \langle\mathbf{p},\mathbf{r}\rangle\ln\left(\frac{1}{\langle\mathbf{p},\mathbf{r}\rangle}\sum_{i=1}^{n} q_i r_i\right)+\sum_{i=1}^{n} r_i p_i\ln\frac{1}{q_i}, \end{aligned}$$

where

$$R_{-\ln}(\mathbf{q},\mathbf{p};\mathbf{r})=\langle\mathbf{q},\mathbf{r}\rangle\ln\left(\frac{1}{\langle\mathbf{q},\mathbf{r}\rangle}\sum_{i=1}^{n}\frac{q_i^2 r_i}{p_i}\right)-\sum_{i=1}^{n} q_i r_i\ln\left(\frac{q_i}{p_i}\right).$$

If in addition $\mathbf{r} = \mathbf{1} = (1,\ldots,1)$ *with* $\sum_{i=1}^{n} q_i = n$, *then*

$$\begin{aligned} &\ln n + \sum_{i=1}^{n} p_i \ln \frac{1}{q_i} - \max_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\} R_{-\ln}(\mathbf{q},\mathbf{p}) \\ \leqslant\ & H(\mathbf{p}) \\ \leqslant\ & \ln n + \sum_{i=1}^{n} p_i \ln \frac{1}{q_i} - \min_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\} R_{-\ln}(\mathbf{q},\mathbf{p}) \\ \leqslant\ & \ln n + \sum_{i=1}^{n} p_i \ln \frac{1}{q_i}, \end{aligned} \tag{40}$$

where

$$R_{-\ln}(\mathbf{q},\mathbf{p}) = n\ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{q_i^2}{p_i}\right) - \sum_{i=1}^{n} q_i \ln\left(\frac{q_i}{p_i}\right).$$

Proof. Applying (37) to the convex function $f(t) = -\ln t$, we get

$$\begin{aligned} 0 &\leqslant \min_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\} R_{-\ln}(\mathbf{q},\mathbf{p};\mathbf{r}) \leqslant \sum_{i=1}^{n} r_i p_i \ln\frac{1}{q_i} - H(\mathbf{p};\mathbf{r}) + \langle \mathbf{p},\mathbf{r}\rangle \ln\left(\frac{1}{\langle \mathbf{p},\mathbf{r}\rangle}\sum_{i=1}^{n} q_i r_i\right) \\ &\leqslant \max_{1\leqslant i\leqslant n}\left\{\frac{p_i}{q_i}\right\} R_{-\ln}(\mathbf{q},\mathbf{p};\mathbf{r}), \end{aligned}$$

which is equivalent to (39).

Further, let $\sum_{i=1}^{n} q_i = n$ and $\mathbf{r} = \mathbf{1} = (1,\ldots,1)$. Then $\langle \mathbf{p},\mathbf{r}\rangle = \sum_{i=1}^{n} p_i = 1$, and from (39) we get (40). □

5. Applications including Zipf-Mandelbrot law

The Zipf-Mandelbrot law is a discrete probability distribution depending on parameters $n \in \mathbb{N}$, $q \geqslant 0$ and $s > 0$ with probability mass function defined with

$$f(k,n,q,s) = \frac{1}{(k+q)^s H_{n,q,s}}, \quad k = 1,2,\ldots,n,$$

where

$$H_{n,q,s} = \sum_{k=1}^{n} \frac{1}{(k+q)^s}. \tag{41}$$

It is also known as the Pareto-Zipf law, a power-law distribution on ranked data, defined by Mandelbrot [23] as generalization of a simpler distribution called Zipf's law [36]. Many naturally phenomena, as earthquake magnitudes, city sizes, incomes, word frequencies and etc., are distributed according to this distribution. It implies that small occurrences are extremely common, whereas large instances are extremely rare. The Zipf-Mandelbrot has wide applications in many branches of science, as well as linguistics, information sciences , ecological field studies and etc.

If we put $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$ in (33), then Shannon's entropy becomes

$$\begin{aligned}\sum_{i=1}^{n} \frac{\ln (i+q)^s H_{n,q,s}}{(i+q)^s H_{n,q,s}} &= \sum_{i=1}^{n} \frac{\ln (i+q)^s}{(i+q)^s H_{n,q,s}} + \sum_{i=1}^{n} \frac{\ln H_{n,q,s}}{(i+q)^s H_{n,q,s}} \\ &= \frac{s}{H_{n,q,s}} \sum_{i=1}^{n} \frac{\ln (i+q)}{(i+q)^s} + \frac{\ln H_{n,q,s}}{H_{n,q,s}} \sum_{i=1}^{n} \frac{1}{(i+q)^s} \\ &= \frac{s}{H_{n,q,s}} \sum_{i=1}^{n} \frac{\ln (i+q)}{(i+q)^s} + \ln H_{n,q,s},\end{aligned}$$

i.e. we get Shannon's entropy for the Zipf-Mandelbrot law, so called the Zipf-Mandelbrot entropy

$$Z(H,q,s) = \frac{s}{H_{n,q,s}} \sum_{k=1}^{n} \frac{\ln(k+q)}{(k+q)^s} + \ln H_{n,q,s}. \tag{42}$$

Applying results from the previous section we will obtain lower and upper bounds for the Zipf-Mandelbrot entropy $Z(H,q,s)$.

COROLLARY 4. *Let $n \in \mathbb{N}$, $q \geqslant 0$, $s > 0$ and $\mathbf{q} = (q_1, \ldots, q_n) \in [0,\infty)^n$ with $\sum_{i=1}^{n} q_i = n$. Let $H_{n,q,s}$ and $Z(H,q,s)$ be defined by (41)-(42), respectively. Then*

$$\begin{aligned}&\ln n + \frac{1}{H_{n,q,s}} \sum_{i=1}^{n} \frac{1}{(i+q)^s} \ln \frac{1}{q_i} - \frac{1}{H_{n,q,s}} \max_{1\leqslant i\leqslant n} \left\{ \frac{1}{(i+q)^s q_i} \right\} S(n,q,s,\mathbf{q}) \\ &\leqslant Z(H,q,s) \\ &\leqslant \ln n + \frac{1}{H_{n,q,s}} \sum_{i=1}^{n} \frac{1}{(i+q)^s} \ln \frac{1}{q_i} - \frac{1}{H_{n,q,s}} \min_{1\leqslant i\leqslant n} \left\{ \frac{1}{(i+q)^s q_i} \right\} S(n,q,s,\mathbf{q}) \\ &\leqslant \ln n + \frac{1}{H_{n,q,s}} \sum_{i=1}^{n} \frac{1}{(i+q)^s} \ln \frac{1}{q_i},\end{aligned} \tag{43}$$

where

$$S(n,q,s,\mathbf{q}) = n \ln \left(\frac{H_{n,q,s}}{n} \sum_{i=1}^{n} (i+q)^s q_i^2 \right) - \sum_{i=1}^{n} q_i \ln \left(q_i (i+q)^s H_{n,q,s} \right).$$

Proof. Applying (40) we get

$$\begin{aligned}&\ln n + \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \ln \frac{1}{q_i} - \max_{1\leqslant i\leqslant n} \left\{ \frac{\frac{1}{(i+q)^s H_{n,q,s}}}{q_i} \right\} S(n,q,s,\mathbf{q}) \\ &\leqslant Z(H,q,s) \\ &\leqslant \ln n + \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \ln \frac{1}{q_i} - \min_{1\leqslant i\leqslant n} \left\{ \frac{\frac{1}{(i+q)^s H_{n,q,s}}}{q_i} \right\} S(n,q,s,\mathbf{q}) \\ &\leqslant \ln n + \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \ln \frac{1}{q_i},\end{aligned}$$

which is equivalent to (43). □

COROLLARY 5. *Let $n \in \mathbb{N}$, $q \geqslant 0$, $s > 0$, $H_{n,q,s}$ and $Z(H,q,s)$ be defined by (41)-(42), respectively. Then*

$$\ln n - \frac{S(n,q,s)}{H_{n,q,s}(1+q)^s} \leqslant Z(H,q,s) \leqslant \ln n - \frac{S(n,q,s)}{H_{n,q,s}(n+q)^s} \leqslant \ln n, \tag{44}$$

where

$$S(n,q,s) = n \ln \left(\frac{H_{n,q,s}}{n} \sum_{i=1}^{n} (i+q)^s \right) - \sum_{i=1}^{n} \ln \left((i+q)^s H_{n,q,s} \right).$$

Proof. If we choose $\mathbf{q} = \mathbf{1} = (1, \dots, 1)$, then (43) becomes

$$\ln n - \frac{1}{H_{n,q,s}} \max_{1 \leqslant i \leqslant n} \left\{ \frac{1}{(i+q)^s} \right\} S(n,q,s) \leqslant Z(H,q,s) \leqslant \ln n - \frac{1}{H_{n,q,s}} \min_{1 \leqslant i \leqslant n} \left\{ \frac{1}{(i+q)^s} \right\} S(n,q,s) \leqslant \ln n,$$

which is equivalent to (44). □

Acknowledgement. We are grateful to the anonymous reviewer who contributed to the improvement of the paper.

This research is partially supported through project KK.01.1.1.02.0027, a project co-financed by the Croatian Government and the European Union through the European Regional Development Fund – the Competitiveness and Cohesion Operational Programme.

REFERENCES

[1] M. ADIL KHAN, S. IVELIĆ BRADANOVIĆ AND J. PEČARIĆ, *Generalizations of Sherman's inequality by Hermite's interpolating polynomial*, Math. Inequal. Appl. 19 (4) (2016), 1181–1192.

[2] M. ADIL KHAN, S. IVELIĆ BRADANOVIĆ AND J. PEČARIĆ, *Generalizations of Sherman's inequality by Hermite's interpolating polynomial and Green function*, Konuralp J. Math. **4** (2) (2016), 255–270.

[3] P. AGARWAL RAVI, S. IVELIĆ BRADANOVIĆ AND J. PEČARIĆ, *Generalizations of Sherman's inequality by Lidstone's interpolating polynomial*, J. Inequal. Appl. **6**, 2016 (2016).

[4] A. BARBIR, S. IVELIĆ BRADANOVIĆ, Đ. PEČARIĆ AND J. PEČARIĆ, *Converse to the Sherman inequality with applications*, Math. Inequal. Appl. **22** (4) (2019), 1405–1419.

[5] M. BELIS, S. GUIAŞU, *A quantitative-qualitative measure of information in cybernetic systems*, IEEE Trans. Information Theory 14 (1968), 593–594.

[6] A. M. BURTEA, *Two examples of weighted majorization*, Annals of the University of Craiova. Mathematics and Computer Science, 37 (2), 2010, 92–99.

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

[8] I. CSISZÁR AND J. KÖRNER, *Information Theory: Coding Theorem for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[9] S. S. DRAGOMIR, *Bounds for the Normalised Jensen Functional*, Bull. Austral. Math. Soc. **74** (2006).

[10] S. S. DRAGOMIR, *Other inequalities for Csiszár divergence and applications*, Preprint, RGMIA Monographs, Victoria University (2000).

[11] S. S. DRAGOMIR, J. E. PEČARIĆ, L. E. PERSSON, *Properties of some functionals related to Jensen's inequality*, Acta Math. Hungar. (70) 1–2 (1996), 129–143.

[12] S. IVELIĆ BRADANOVIĆ, *More Accurate Majorization Inequalities Obtained Via Superquadraticity and Convexity with Application to Entropies*, Mediterr. J. Math., 18 (2021), 2021; 1–16.

[13] S. IVELIĆ BRADANOVIĆ, *Sherman's inequality and its converse for strongly convex functions with applications to generalized f-divergences*, Turk. J. Math., **43** (2019), 6; 2680–2696.

[14] S. IVELIĆ BRADANOVIĆ, N. LATIF AND J. PEČARIĆ, *On an upper bound for Sherman's inequality*, J. Inequal. Appl. 2016 (2016).

[15] S. IVELIĆ BRADANOVIĆ, N. LATIF, Đ. PEČARIĆ AND J. PEČARIĆ, *Sherman's and related inequalities with applications in information theory*, J. Inequal. Appl. 2018 (2018).

[16] S. IVELIĆ BRADANOVIĆ, J. MIĆIĆ AND J. PEČARIĆ, *Sherman's operator inequality*, J. Math. Inequal. **15** (2021).

[17] S. IVELIĆ BRADANOVIĆ AND J. PEČARIĆ, *Extensions and improvements of Sherman's and related inequalities for n-convex functions*, Open Math. **15** (1) 2017.

[18] S. IVELIĆ BRADANOVIĆ AND J. PEČARIĆ, *Generalizations of Sherman's inequality*, Per. Math. Hung. **74** (2) 2017.

[19] P. A. KLUZA AND M. NIEZGODA, *On Csiszár and Tsallis type f-divergences induced by superquadratic and convex functions*, Math. Inequal. Appl. **21** (2) (2018) 455–467.

[20] M. KRNIĆ, N. LOVRIČEVIĆ, J. PEČARIĆ AND J. PERIĆ, *Superadditivity and monotonicity of the Jensen-type functionals (New methods for improving the Jensen-type inequalities in real and in operator cases)*, Zagreb, Element, 2015.

[21] S. KULLBACK, R. A. LEIBLER, *On information and sufficiency*, The Annals of Mathematical Statistics 22 (1) (1951) 79–86.

[22] D. S. MITRINOVIĆ, J. E. PEČARIĆ, A. M. FINK, *Classical and New Inequalities in Analysis*, Dortrecht-Boston-London: Kluwer Acad. Publ., 1993.

[23] B. MANDELBROT, *Information theory and psycholinguistics: a theory of words frequencies*, in: P. Lazafeld, N. Henry (Eds.), Readings in Mathematical Social. Science, (MIT Press, Cambridge, MA, 1966).

[24] M. MOHAN, J. MITTER, *On bounds of 'useful' information measure*, Indian J. Pure Appl. Math. **9** (9), 960–964 (1978).

[25] C. P. NICULESCU AND I. ROVENŢA, *An approach of majorization in spaces with a curved geometry*, J. Math. Anal. Appl. **411** (2014), 119–128.

[26] C. P. NICULESCU AND I. ROVENŢA, *Relative convexity and its applications*, Aequationes Math. **89** (2015), 1389–1400.

[27] M. NIEZGODA, *Nonlinear Sherman-type inequalities*, Adv. Nonlinear Anal. **9** (1) (2020), 168–175.

[28] M. NIEZGODA, *Remarks on convex functions and separable sequences*, Discrete Math. 308 (2008), 1765–1773.

[29] M. NIEZGODA, *Remarks on Sherman like inequalities for (α,β)-convex functions*, Math. Ineqal. Appl. **17** (4) (2014), 1579–1590.

[30] M. NIEZGODA, *Sherman, Hermite-Hadamard and Fejér like Inequalities for Convex Sequences and Nondecreasing Convex Functions*, Filomat **31** (8) (2017), pp. 2321–2335.

[31] M. NIEZGODA, *Vector joint majorization and generalization of Csiszar-Korner's inequality for f-divergence*, Discrete Appl. Math., 198 (2016), 195–205.

[32] J. PEČARIĆ, F. PROSCHAN AND Y. L. TONG, *Convex functions, Partial Orderings and Statistical Applications*, Academic Press, New York, 1992.

[33] C. E. SHANNON AND W. WEAVER, *The Mathemtiatical Theory of Comnunication*, Urbana, University of Illinois Press, 1949.

[34] S. SHERMAN, *On a theorem of Hardy, Littlewood, Pólya and Blackwell*, Proc. Nat. Acad. Sci. USA, **37** (1) (1957), 826–831.

[35] A. SRIVASTAVA, *Some new bounds of weighted entropy measures*, Cybern. Inf. Technol. 11 (3), 60–65.

[36] G. K. ZIPF, *Selected Studies of the Principle of Relative Frequency in Language*, Cambridge, MA: Harvard University Press. 1932.

(Received March 16, 2022)

Slavica Ivelić Bradanović
Faculty of Civil Engineering, Architecture And Geodesy
University of Split
Matice Hrvatske 15, 21000 Split, Croatia
e-mail: sivelic@gradst.hr

Đilda Pečarić
University North
Trg dr. Žarka Dolinara 1, 48000 Koprivnica, Croatia
e-mail: djilda.pecaric@unin.hr

Josip Pečarić
Croatian Academy of Sciences and Arts
Trg Nikole Šubića Zrinskog 11, 10000 Zagreb, Croatia
e-mail: pecaric@element.hr