

RESEARCH

Open Access



Sherman's and related inequalities with applications in information theory

S. Ivelić Bradanović¹, N. Latif^{2*} , Đ. Pečarić³ and J. Pečarić^{4,5}

*Correspondence:
naveed707@gmail.com
²Department of General Studies,
Jubail Industrial College, Jubail
Industrial City, Kingdom of Saudi
Arabia
Full list of author information is
available at the end of the article

Abstract

In this paper we give extensions of Sherman's inequality considering the class of convex functions of higher order. As particular cases, we get an extended weighted majorization inequality as well as Jensen's inequality which have direct connection to information theory. We use the obtained results to derive new estimates for Shannon's and Rényi's entropy, information energy, and some well-known measures between probability distributions. Using the Zipf–Mandelbrot law, we introduce new functionals to derive some related results.

MSC: 26D15; 26D20; 26D99; 94A17

Keywords: Sherman theorem; Majorization inequality; Jensen inequality; Green function; Abel–Gontscharoff interpolating polynomial; n -convex function; Entropy; Information theory; ϕ -divergence; Zipf–Mandelbrot law

1 Introduction and preliminaries

We start with a brief overview of divided differences and n -convex functions and give some basic results from the majorization theory.

An n th order divided difference of a function $\phi : [\alpha, \beta] \rightarrow \mathbb{R}$ at distinct points $x_0, x_1, \dots, x_n \in [\alpha, \beta]$ may be defined recursively by

$$[x_i; \phi] = \phi(x_i), \quad i = 0, \dots, n,$$
$$[x_0, \dots, x_n; \phi] = \frac{[x_1, \dots, x_n; \phi] - [x_0, \dots, x_{n-1}; \phi]}{x_n - x_0}.$$

The value $[x_0, \dots, x_n; \phi]$ is independent of the order of the points x_0, \dots, x_n .

A function ϕ is n -convex on $[\alpha, \beta]$ if

$$[x_0, x_1, \dots, x_n; \phi] \geq 0$$

holds for all choices of $(n + 1)$ distinct points $x_i \in [\alpha, \beta]$, $i = 0, \dots, n$.

Remark 1 From this definition it follows that 1-convex function is an increasing function and 2-convex function is just a convex function. If $\phi^{(n)}$ exists, then ϕ is n -convex iff $\phi^{(n)} \geq 0$. Also, if ϕ is n -convex for $n \geq 2$, then $\phi^{(k)}$ exists and ϕ is $(n - k)$ -convex for $1 \leq k \leq n - 2$. For more information, see [1].

For two vectors $\mathbf{x}, \mathbf{y} \in [\alpha, \beta]^l$, let $x_{[i]}, y_{[i]}$ denote the i th largest entries of \mathbf{x} and \mathbf{y} , respectively. It is well known that

$$\sum_{i=1}^m y_{[i]} \leq \sum_{i=1}^m x_{[i]} \quad \text{for } m = 1, 2, \dots, l-1 \quad \text{and} \quad \sum_{i=1}^l x_i = \sum_{i=1}^l y_i,$$

i.e., we say that \mathbf{x} majorizes \mathbf{y} , in symbol $\mathbf{y} \prec \mathbf{x}$, iff

$$\mathbf{y} = \mathbf{x}\mathbf{A}$$

for some doubly stochastic matrix $\mathbf{A} = (a_{ij}) \in \mathcal{M}_l(\mathbb{R})$, i.e., a matrix with nonnegative entries and rows and columns sums equal to 1. Moreover, $\mathbf{y} \prec \mathbf{x}$ implies

$$\sum_{i=1}^l \phi(y_i) \leq \sum_{i=1}^l \phi(x_i)$$

for every continuous convex function $\phi : [\alpha, \beta] \rightarrow \mathbb{R}$. This result, obtained by Hardy et al. (1929 [2]), is well known as a majorization inequality and plays an important role in the study of majorization theory.

Sherman [3] considered the weighted concept of majorization

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

between two vectors $\mathbf{x} = (x_1, \dots, x_m) \in [\alpha, \beta]^m$ and $\mathbf{y} = (y_1, \dots, y_l) \in [\alpha, \beta]^l$ with nonnegative weights $\mathbf{a} = (a_1, \dots, a_m)$ and $\mathbf{b} = (b_1, \dots, b_l)$. The concept of weighted majorization is defined by the assumption of existence of a matrix $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{lm}(\mathbb{R})$ such that

$$a_{ij} \geq 0 \quad \text{for all } i, j, \tag{1.1}$$

$$\sum_{j=1}^m a_{ij} = 1, \quad i = 1, \dots, l, \tag{1.2}$$

$$a_j = \sum_{i=1}^l b_i a_{ij}, \quad j = 1, \dots, m, \tag{1.3}$$

$$y_i = \sum_{j=1}^m x_j a_{ij}, \quad i = 1, \dots, l. \tag{1.4}$$

The matrix $\mathbf{A} = (a_{ij}) \in \mathcal{M}_{lm}(\mathbb{R})$ with conditions (1.1) and (1.2) is called row stochastic matrix. Sherman proved that under conditions (1.1)–(1.4) for every convex function $\phi : [\alpha, \beta] \rightarrow \mathbb{R}$, the inequality

$$\sum_{i=1}^l b_i \phi(y_i) \leq \sum_{j=1}^m a_j \phi(x_j) \tag{1.5}$$

holds. We can write conditions (1.3) and (1.4) in the form

$$\mathbf{a} = \mathbf{b}\mathbf{A} \quad \text{and} \quad \mathbf{y} = \mathbf{x}\mathbf{A}^T, \tag{1.6}$$

where \mathbf{A}^T denotes the transpose matrix.

As a special case of Sherman's inequality, when $l = m$ and $a_j = b_i$, for all $i, j = 1, \dots, m$, we get the weighted version of majorization inequality

$$\sum_{i=1}^m a_i \phi(y_i) \leq \sum_{i=1}^m a_i \phi(x_i).$$

Putting $\sum_{i=1}^m a_i = 1$ and $y_1 = y_2 = \dots = y_m = \sum_{i=1}^m a_i x_i$, we get Jensen's inequality in the form

$$\phi\left(\sum_{i=1}^m a_i x_i\right) \leq \sum_{i=1}^m a_i \phi(x_i). \quad (1.7)$$

We can get Jensen's inequality (1.7) directly from (1.5) by setting $l = 1$ and $\mathbf{b} = (1)$.

The concept of majorization has a large number of appearances in many different fields of applications, particular in many branches of mathematics. A complete and superb reference on the subject is the monograph [4], and many results from the theory of majorization are directly or indirectly inspired by it. In this paper we give extensions of Sherman's inequality by considering the class of convex functions of higher order. As a particular case, we get an extension of weighted majorization inequality and Jensen's inequality which can be used to derive some new estimates for some entropies and measures between probability distributions. Also, we use the Zipf–Mandelbrot law to illustrate the obtained results.

2 Some technical lemmas

In this section we present two technical lemmas that give us two identities which will be very useful for us to obtain main results.

Let us consider the function $G : [\alpha, \beta] \times [\alpha, \beta] \rightarrow \mathbb{R}$ defined by

$$G(x, y) = \begin{cases} \frac{(x-\beta)(y-\alpha)}{\beta-\alpha}, & \alpha \leq y \leq x, \\ \frac{(y-\beta)(x-\alpha)}{\beta-\alpha}, & x \leq y \leq \beta, \end{cases} \quad (2.1)$$

which presents Green's function of the boundary value problem

$$z'' = 0, \quad z(\alpha) = z(\beta) = 0.$$

This function is convex and continuous with respect to both variables x and y .

Integration by parts easily yields that, for any function $\phi \in C^2([\alpha, \beta])$, the following holds:

$$\phi(t) = \frac{\beta-t}{\beta-\alpha} \phi(\alpha) + \frac{t-\alpha}{\beta-\alpha} \phi(\beta) + \int_{\alpha}^{\beta} G(t, y) \phi''(y) dy. \quad (2.2)$$

Applying (2.2) to Sherman's difference $\sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i)$, we obtain the first identity.

Lemma 1 *Let $\mathbf{x} \in [\alpha, \beta]^m$, $\mathbf{y} \in [\alpha, \beta]^l$, $\mathbf{a} \in \mathbb{R}^m$, and $\mathbf{b} \in \mathbb{R}^l$ be such that (1.6) holds for some matrix $\mathbf{A} \in \mathcal{M}_{lm}(\mathbb{R})$ with $\sum_{j=1}^m a_{ij} = 1$, $i = 1, \dots, l$. Let G be defined by (2.1). Then, for every*

function $\phi \in C^2([\alpha, \beta])$, the following identity holds:

$$\sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i) = \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) \phi''(y) dy. \tag{2.3}$$

Proof Using (2.2) in Sherman’s difference, we have

$$\begin{aligned} & \sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i) \\ &= \frac{\beta \phi(\alpha) - \alpha \phi(\beta)}{\beta - \alpha} \left(\sum_{j=1}^m a_j - \sum_{i=1}^l b_i \right) + \frac{\phi(\beta) - \phi(\alpha)}{\beta - \alpha} \left(\sum_{j=1}^m a_j x_j - \sum_{i=1}^l b_i y_i \right) \\ & \quad + \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) \phi''(y) dy. \end{aligned}$$

Since (1.3) and (1.4) hold, then we have

$$\sum_{j=1}^m a_j - \sum_{i=1}^l b_i = 0 \quad \text{and} \quad \sum_{j=1}^m a_j x_j - \sum_{i=1}^l b_i y_i = 0,$$

i.e., we get identity (2.3). □

We use the Abel–Gontscharoff interpolation for two points with integral remainder to obtain another identity.

Let $n, k \in \mathbb{N}$, $n \geq 2$, $0 \leq k \leq n - 1$, and $\phi \in C^n([\alpha, \beta])$. Then

$$\phi(t) = P_{AG}(t) + e_{AG}(t), \tag{2.4}$$

where

$$P_{AG}(t) = \sum_{s=0}^k \frac{(t - \alpha)^s}{s!} \phi^{(s)}(\alpha) + \sum_{r=0}^{n-k-2} \left[\sum_{s=0}^r \frac{(t - \alpha)^{k+1+s} (\alpha - \beta)^{r-s}}{(k + 1 + s)! (r - s)!} \right] \phi^{(k+1+r)}(\beta)$$

is the Abel–Gontscharoff interpolating polynomial for two points of degree $n - 1$, and the remainder is given by

$$e_{AG}(t) = \int_{\alpha}^{\beta} G_n(t, u) \phi^{(n)}(u) du,$$

where

$$G_{n,k}(t, u) = \frac{1}{(n - 1)!} \begin{cases} \sum_{s=0}^k \binom{n-1}{s} (t - \alpha)^s (\alpha - u)^{n-s-1}, & \alpha \leq u \leq t; \\ -\sum_{s=k+1}^{n-1} \binom{n-1}{s} (t - \alpha)^s (\alpha - u)^{n-s-1}, & t \leq u \leq \beta. \end{cases} \tag{2.5}$$

Further, for $\alpha \leq u, t \leq \beta$, the following inequalities hold:

$$\begin{aligned}
 (-1)^{n-k-1} \frac{\partial^s G_{n,k}(t, u)}{\partial t^s} &\geq 0, \quad 0 \leq s \leq k, \\
 (-1)^{n-s} \frac{\partial^s G_{n,k}(t, u)}{\partial t^s} &\geq 0, \quad k+1 \leq s \leq n-1.
 \end{aligned}
 \tag{2.6}$$

For more information, see [5].

Now we use interpolation (2.4) on ϕ'' to obtain the second identity.

Lemma 2 *Let $\mathbf{x} \in [\alpha, \beta]^m, \mathbf{y} \in [\alpha, \beta]^l, \mathbf{a} \in \mathbb{R}^m$, and $\mathbf{b} \in \mathbb{R}^l$ be such that (1.6) holds for some matrix $\mathbf{A} \in \mathcal{M}_{lm}(\mathbb{R})$ with $\sum_{j=1}^m a_{ij} = 1, i = 1, \dots, l$. Let $n, k \in \mathbb{N}, n \geq 4, 0 \leq k \leq n - 1$, and $G, G_{n,k}$ be defined by (2.1), (2.5), respectively. Then, for every function $\phi \in C^n([\alpha, \beta])$, the following identity holds:*

$$\begin{aligned}
 &\sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i) \\
 &= \sum_{s=0}^k \frac{\phi^{(s+2)}(\alpha)}{s!} \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (y - \alpha)^s dy \\
 &\quad + \sum_{r=0}^{n-k-4} \sum_{s=0}^r \frac{(-1)^{r-s} (\beta - \alpha)^{r-s}}{(k+1+s)!(r-s)!} \phi^{(k+3+r)}(\beta) \\
 &\quad \times \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (y - \alpha)^{k+1+s} dy \\
 &\quad + \int_{\alpha}^{\beta} \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) G_{n-2,k}(y, u) \phi^{(n)}(u) du dy.
 \end{aligned}
 \tag{2.7}$$

Proof If we apply formula (2.4) to a function ϕ'' , it implies substitution of n with $n - 2$ in (2.4), and we get

$$\begin{aligned}
 &\phi''(y) \\
 &= \sum_{s=0}^k \frac{(y - \alpha)^s}{s!} \phi^{(s+2)}(\alpha) + \sum_{r=0}^{n-k-4} \left[\sum_{s=0}^r \frac{(y - \alpha)^{k+1+s} (\alpha - \beta)^{r-s}}{(k+1+s)!(r-s)!} \right] \phi^{(k+3+r)}(\beta) \\
 &\quad + \int_{\alpha}^{\beta} G_{n-2,k}(y, u) \phi^{(n)}(u) du.
 \end{aligned}
 \tag{2.8}$$

Using (2.8) in (2.3), we obtain the required result. □

3 Extensions of Sherman’s inequality

We start this section with an extension of Sherman’s inequality to a more general class of n -convex functions.

Theorem 1 *Let $\mathbf{x} \in [\alpha, \beta]^m, \mathbf{y} \in [\alpha, \beta]^l, \mathbf{a} \in \mathbb{R}^m$, and $\mathbf{b} \in \mathbb{R}^l$ be such that (1.6) holds for some matrix $\mathbf{A} \in \mathcal{M}_{lm}(\mathbb{R})$ with $\sum_{j=1}^m a_{ij} = 1, i = 1, \dots, l$. Let $n, k \in \mathbb{N}, n \geq 4, 0 \leq k \leq n - 1$,*

and $G, G_{n,k}$ be defined by (2.1), (2.5), respectively. If $\phi \in C^n([\alpha, \beta])$ is n -convex and

$$\int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) G_{n-2,k}(y, u) dy \geq 0, \quad \text{for all } u \in [\alpha, \beta], \tag{3.1}$$

then

$$\begin{aligned} & \sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i) \\ & \geq \sum_{s=0}^k \frac{\phi^{(s+2)}(\alpha)}{s!} \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (y - \alpha)^s dy \\ & \quad + \sum_{r=0}^{n-k-4} \sum_{s=0}^r \frac{(-1)^{r-s} (\beta - \alpha)^{r-s}}{(k + 1 + s)!(r - s)!} \phi^{(k+3+r)}(\beta) \\ & \quad \times \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (y - \alpha)^{k+1+s} dy. \end{aligned} \tag{3.2}$$

If the reverse inequality in (3.1) holds, then also the reverse inequality in (3.2) holds.

Proof Under the assumptions of the theorem, identity (2.7) holds. Since ϕ is n -convex, then $\phi^{(n)} \geq 0$ on $[\alpha, \beta]$. Therefore, if (3.1) is satisfied, then inequality (3.2) holds. \square

Remark 2 Since we have $(-1)^{n-k-3} G_{n-2,k}(y, u) \geq 0$ by (2.6), then in case $n - k$ is odd, instead assumption (3.1), it is enough to assume that

$$\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \geq 0, \quad \text{for } y \in [\alpha, \beta].$$

The following extension of Sherman’s inequality, under Sherman’s condition of nonnegativity of vectors \mathbf{a}, \mathbf{b} , and matrix \mathbf{A} , also holds.

Theorem 2 Let $\mathbf{x} \in [\alpha, \beta]^m, \mathbf{y} \in [\alpha, \beta]^l, \mathbf{a} \in [0, \infty)^m$, and $\mathbf{b} \in [0, \infty)^l$ be such that (1.6) holds for some row stochastic matrix $\mathbf{A} \in \mathcal{M}_{lm}(\mathbb{R})$. Let $n, k \in \mathbb{N}, n \geq 4, 0 \leq k \leq n - 1$, be such that $n - k$ is odd. Let $G, G_{n,k}$ be defined by (2.1), (2.5), respectively, and $\phi \in C^n([\alpha, \beta])$ be n -convex. Then inequality (3.2) holds.

Proof Since by (2.6) we have $(-1)^{n-k-3} G_{n-2,k}(y, u) \geq 0$, then, when $n - k$ is odd, we have

$$G_{n-2,k}(y, u) \geq 0.$$

Further, $G(\cdot, y), y \in [\alpha, \beta]$, is convex on $[\alpha, \beta]$, and by Sherman’s inequality, we have

$$\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \geq 0.$$

Combining these two facts, assumption (3.1) is satisfied. Hence by Theorem 1, inequality (3.2) holds. □

Remark 3 In case $n - k$ is even, then the reverse inequality in (3.1) holds, i.e., the reverse inequality in (3.2) holds.

Theorem 3 *Let all the assumptions of Theorem 2 be satisfied.*

- (i) *If $\phi^{(s+2)}(\alpha) \geq 0$ for each $s = 0, \dots, k$ and $(-1)^{r-s}\phi^{(k+3+r)}(\beta) \geq 0$ for each $r = 0, \dots, n - k - 4$ and $s = 0, \dots, r$, then*

$$\begin{aligned} & \sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i) \\ & \geq \sum_{s=0}^k \frac{\phi^{(s+2)}(\alpha)}{s!} \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (y - \alpha)^s dy \\ & \quad + \sum_{r=0}^{n-k-4} \sum_{s=0}^r \frac{(-1)^{r-s}(\beta - \alpha)^{r-s}}{(k + 1 + s)!(r - s)!} \phi^{(k+3+r)}(\beta) \\ & \quad \times \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (y - \alpha)^{k+1+s} dy \\ & \geq 0. \end{aligned} \tag{3.3}$$

- (ii) *If the function*

$$\begin{aligned} F(\cdot) = & \sum_{s=0}^k \frac{\phi^{(s+2)}(\alpha)}{s!} \int_{\alpha}^{\beta} G(\cdot, y) (y - \alpha)^s dy \\ & + \sum_{r=0}^{n-k-4} \sum_{s=0}^r \frac{(-1)^{r-s}(\beta - \alpha)^{r-s}}{(k + 1 + s)!(r - s)!} \phi^{(k+3+r)}(\beta) \int_{\alpha}^{\beta} G(\cdot, y) (y - \alpha)^{k+1+s} dy \end{aligned} \tag{3.4}$$

is convex on $[\alpha, \beta]$, then (3.3) holds.

Proof (i) Under the assumptions, the nonnegativity of the right-hand side of (3.2) is obvious, i.e., the double inequality (3.3) holds.

(ii) The right-hand side of (3.2) can be written in the form $\sum_{j=1}^m a_j F(x_j) - \sum_{i=1}^l b_i F(y_i)$. So, if F is convex, then by Sherman’s inequality we have

$$\sum_{j=1}^m a_j F(x_j) - \sum_{i=1}^l b_i F(y_i) \geq 0,$$

i.e., we again get the nonnegativity of the right-hand side of (3.2), which we need to prove. □

Remark 4 Note that inequality (3.3) includes a new lower bound for Sherman’s difference in the form

$$\begin{aligned}
 A_n(\phi; \alpha, \beta) &= \sum_{s=0}^k \frac{\phi^{(s+2)}(\alpha)}{s!} \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (y - \alpha)^s dy \\
 &\quad + \sum_{r=0}^{n-k-4} \sum_{s=0}^r \frac{(-1)^{r-s} (\beta - \alpha)^{r-s}}{(k + 1 + s)!(r - s)!} \phi^{(k+3+r)}(\beta) \\
 &\quad \times \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (y - \alpha)^{k+1+s} dy. \tag{3.5}
 \end{aligned}$$

Specially, for $n = 4, k = 1$, the lower bound has the form

$$\int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) (\phi''(\alpha) + \phi'''(\alpha)(y - \alpha)) dy. \tag{3.6}$$

Using notation $\| \cdot \|_p$ for the standard p -norm and applying the well-known Hölder inequality, we obtain the following result.

Theorem 4 Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $\mathbf{x} \in [\alpha, \beta]^m, \mathbf{y} \in [\alpha, \beta]^l, \mathbf{a} \in \mathbb{R}^m$, and $\mathbf{b} \in \mathbb{R}^l$ be such that (1.6) holds for some matrix $\mathbf{A} \in \mathcal{M}_{lm}(\mathbb{R})$ with $\sum_{j=1}^m a_{ij} = 1, i = 1, \dots, l$. Let $n, k \in \mathbb{N}, n \geq 4, 0 \leq k \leq n - 1, \phi \in C^n([\alpha, \beta])$, and $G, G_{n,k}$ be defined by (2.1), (2.5), respectively. Then

$$\begin{aligned}
 &\left| \sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i) - A_n(\phi; \alpha, \beta) \right| \\
 &\leq \|\phi^{(n)}\|_p \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) G_{n-2,k}(y, u) dy \right)^q du \right)^{\frac{1}{q}},
 \end{aligned}$$

where $A_n(\phi; \alpha, \beta)$ is defined by (3.5).

Proof Under the assumptions of the theorem, identity (2.7) holds. Applying Hölder’s inequality to (2.7), we get

$$\begin{aligned}
 &\left| \sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i) - A_n(\phi; \alpha, \beta) \right| \\
 &= \left| \int_{\alpha}^{\beta} \int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) G_{n-2,k}(y, u) \phi^{(n)}(u) du dy \right| \\
 &\leq \|\phi^{(n)}\|_p \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \left(\sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y) \right) G_{n-2,k}(y, u) dy \right)^q du \right)^{\frac{1}{q}}. \quad \square
 \end{aligned}$$

As a direct consequence of the previous results, choosing $n = 4$ and $k = 1$, we get the following corollary.

Corollary 1 Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let G be defined by (2.1), $\mathbf{x} \in [\alpha, \beta]^m, \mathbf{y} \in [\alpha, \beta]^l, \mathbf{a} \in [0, \infty)^m$, and $\mathbf{b} \in [0, \infty)^l$ be such that (1.6) holds for some row stochastic matrix $\mathbf{A} \in \mathcal{M}_{lm}(\mathbb{R})$. If $\phi \in C^4([\alpha, \beta])$ is 4-convex, then

$$\begin{aligned} 0 &\leq \sum_{j=1}^m a_j \phi(x_j) - \sum_{i=1}^l b_i \phi(y_i) - \int_{\alpha}^{\beta} \mathcal{G}(y) (\phi''(\alpha) + \phi'''(\alpha)(y - \alpha)) dy \\ &\leq \|\phi^{(4)}\|_p \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{3.7}$$

where $\mathcal{G}(y) = \sum_{j=1}^m a_j G(x_j, y) - \sum_{i=1}^l b_i G(y_i, y)$ and

$$G_{2,1}(y, u) = \begin{cases} y - u, & \alpha \leq y \leq u, \\ 0, & u \leq y \leq \beta. \end{cases} \tag{3.8}$$

Remark 5 Specially, if we set $l = m$ and $a_j = b_i$ for each $i, j = 1, \dots, l$, from the previous result, as a direct consequence, we obtain the following extension of majorization inequality:

$$\begin{aligned} 0 &\leq \sum_{i=1}^m a_i \phi(x_i) - \sum_{i=1}^m a_i \phi(y_i) - \int_{\alpha}^{\beta} \mathcal{G}(y) (\phi''(\alpha) + \phi'''(\alpha)(y - \alpha)) dy \\ &\leq \|\phi^{(4)}\|_p \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{3.9}$$

where $\mathcal{G}(y) = \sum_{i=1}^m a_i G(x_i, y) - \sum_{i=1}^m a_i G(y_i, y)$.

Remark 6 By setting $l = 1, \mathbf{b} = (1)$, from (3.7), as a direct consequence, we get the extension of Jensen's inequality

$$\begin{aligned} 0 &\leq \sum_{i=1}^m a_i \phi(x_i) - \phi \left(\sum_{i=1}^m a_i x_i \right) - \int_{\alpha}^{\beta} \mathcal{G}(y) (\phi''(\alpha) + \phi'''(\alpha)(y - \alpha)) dy \\ &\leq \|\phi^{(4)}\|_p \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{3.10}$$

where $\mathcal{G}(y) = \sum_{i=1}^m a_i G(x_i, y) - G(\sum_{i=1}^m a_i x_i, y)$.

4 Applications in information theory

Throughout the rest of paper, let α, β be positive real numbers such that $0 < \alpha < \beta$.

By X we denote a discrete random variable with distribution

$$\begin{pmatrix} x_1 & x_2 & \dots & x_m \\ p_1 & p_2 & \dots & p_m \end{pmatrix},$$

where $\mathbf{p} = (p_1, \dots, p_m)$ is a positive probability distribution, i.e., $p_i > 0, i = 1, \dots, m$, with $\sum_{i=1}^m p_i = 1$.

Shannon entropy [6] is defined by

$$H(X) = \sum_{i=1}^m p_i \ln \frac{1}{p_i}.$$

It is well known that the maximum possible value of $H(X)$ concerns in terms of the size of $R(X)$, i.e., the inequality

$$0 \leq H(X) \leq \ln m$$

holds. Furthermore, $H(X) = 0$ iff $p_i = 1$ for some i , and $H(X) = \ln m$ iff $p_i = \frac{1}{m}$ for all $i = 1, \dots, m$. Some related results can be found in [7–13].

Corollary 2 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let $\xi = (\xi_1, \dots, \xi_m) \in [\alpha, \beta]^m$ and $\mathbf{p} = (p_1, \dots, p_m)$ be a positive probability distribution. Then*

$$\begin{aligned} 0 &\leq \ln \left(\sum_{i=1}^m p_i \xi_i \right) - \sum_{i=1}^m p_i \ln \xi_i - \int_{\alpha}^{\beta} \mathcal{G}(y) \frac{3\alpha - 2y}{\alpha^3} dy \\ &\leq 6 \left(\frac{\beta^{1-4p} - \alpha^{1-4p}}{1 - 4p} \right)^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{4.1}$$

where $\mathcal{G}(y) = \sum_{i=1}^m p_i G(\xi_i, y) - G(\sum_{i=1}^m p_i \xi_i, y)$.

Proof Substituting ξ_i in place of x_i, p_i in place of a_i in (3.10) and choosing $\phi(x) = -\ln x$, we obtain (4.1). □

Corollary 3 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let \mathbf{p} be a positive probability distribution with $p_i^{-1} \in [\alpha, \beta], i = 1, \dots, m$. Then*

$$\begin{aligned} 0 &\leq \ln m - H(X) - \int_{\alpha}^{\beta} \mathcal{G}(y) \frac{3\alpha - 2y}{\alpha^3} dy \\ &\leq 6 \left(\frac{\beta^{1-4p} - \alpha^{1-4p}}{1 - 4p} \right)^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned}$$

where $\mathcal{G}(y) = \sum_{i=1}^m p_i G(p_i^{-1}, y) - \sum_{i=1}^m G(m, y)$.

Proof If we substitute $\frac{1}{p_i}$ in place of ξ_i in (4.1), we get the required result. □

Rényi's entropy [14] of order $\lambda, \lambda \in (0, 1) \cup (1, \infty)$, is defined by

$$H_{\lambda}(X) = \frac{1}{1 - \lambda} \ln \left(\sum_{i=1}^m p_i^{\lambda} \right).$$

Applying discrete Jensen’s inequality to the convex function $\phi(x) = -\ln x$, we have

$$\ln\left(\sum_{i=1}^m p_i x_i\right) \geq \sum_{i=1}^m p_i \ln x_i.$$

Substituting $p_i^{\lambda-1}$ in place of x_i , we get

$$\ln\left(\sum_{i=1}^m p_i^\lambda\right) \geq (\lambda - 1) \sum_{i=1}^m p_i \ln p_i,$$

which is equivalent to

$$(1 - \lambda)[H_\lambda(X) - H(X)] \geq 0.$$

Specially, we have

$$H_\lambda(X) \geq H(X), \quad \lambda \in (0, 1),$$

$$H_\lambda(X) \leq H(X), \quad \lambda \in (1, \infty)$$

with the equality in case of the uniform distribution, i.e., when $p_i = \frac{1}{m}, i = 1, \dots, m$.

Corollary 4 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let $\lambda \in (0, 1) \cup (1, \infty)$ and \mathbf{p} be a positive probability distribution with $p_i^{\lambda-1} \in [\alpha, \beta], i = 1, \dots, m$. Then*

$$\begin{aligned} 0 &\leq (1 - \lambda)[H_\lambda(X) - H(X)] - \int_\alpha^\beta \mathcal{G}(y) \frac{3\alpha - 2y}{\alpha^3} dy \\ &\leq 6 \left(\frac{\beta^{1-4p} - \alpha^{1-4p}}{1 - 4p}\right)^{\frac{1}{p}} \left(\int_\alpha^\beta \left(\int_\alpha^\beta \mathcal{G}(y) G_{2,1}(y, u) dy\right)^q du\right)^{\frac{1}{q}}, \end{aligned}$$

where $\mathcal{G}(y) = \sum_{i=1}^m p_i G(p_i^{\lambda-1}, y) - G(\sum_{i=1}^m p_i^\lambda, y)$.

Proof Substituting $p_i^{\lambda-1}$ in place of ξ_i in (4.1), we obtain the required result. □

The information energy of the random variable X is defined by

$$E(X) = \sum_{i=1}^m p_i^2.$$

Corollary 5 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let $\xi = (\xi_1, \dots, \xi_m) \in [\alpha, \beta]^m$ and $\mathbf{p} = p_1, \dots, p_m$ be a positive probability distribution.*

(i) *For $\lambda \in (0, 1)$, we have*

$$0 \leq \left(\sum_{i=1}^m p_i \xi_i\right)^\lambda - \sum_{i=1}^m p_i \xi_i^\lambda$$

$$\begin{aligned}
 & - \int_{\alpha}^{\beta} \mathcal{G}(y) (\lambda(1-\lambda)\alpha^{\lambda-2} - \lambda(1-\lambda)(2-\lambda)\alpha^{\lambda-3}(y-\alpha)) dy \\
 & \leq \lambda(1-\lambda)(2-\lambda)(3-\lambda) \left(\frac{\beta^{(\lambda-4)p+1} - \alpha^{(\lambda-4)p+1}}{(\lambda-4)p+1} \right)^{\frac{1}{p}} \\
 & \quad \times \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y,u) dy \right)^q du \right)^{\frac{1}{q}}, \tag{4.2}
 \end{aligned}$$

where

$$\mathcal{G}(y) = \sum_{i=1}^m p_i G(\xi_i, y) - G\left(\sum_{i=1}^m p_i \xi_i, y\right). \tag{4.3}$$

(ii) For $\lambda \in (1, \infty)$, we have

$$\begin{aligned}
 0 & \leq \sum_{i=1}^m p_i \xi_i^{\lambda} - \left(\sum_{i=1}^m p_i \xi_i \right)^{\lambda} \\
 & \quad - \int_{\alpha}^{\beta} \mathcal{G}(y) (\lambda(\lambda-1)\alpha^{\lambda-2} + \lambda(\lambda-1)(\lambda-2)\alpha^{\lambda-3}(y-\alpha)) dy \\
 & \leq \lambda(\lambda-1)|(\lambda-2)(\lambda-3)| \left(\frac{\beta^{(\lambda-4)p+1} - \alpha^{(\lambda-4)p+1}}{(\lambda-4)p+1} \right)^{\frac{1}{p}} \\
 & \quad \times \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y,u) dy \right)^q du \right)^{\frac{1}{q}},
 \end{aligned}$$

where $\mathcal{G}(y)$ is defined by (4.3).

Proof (i) Substituting ξ_i in place of x_i , p_i in place of a_i in (3.10), and choosing $\phi(x) = -x^{\lambda}$, $\lambda \in (0, 1)$, we obtain (4.2).

(ii) Substituting ξ_i in place of x_i , p_i in place of a_i in (3.10), and choosing $\phi(x) = x^{\lambda}$, $\lambda \in (1, \infty)$, we obtain the required result. □

Corollary 6 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty$, $1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let $\mathbf{p} \in [\alpha, \beta]^m$ be a positive probability distribution.*

(i) For $\lambda \in (0, 1)$, we have

$$\begin{aligned}
 0 & \leq E^{\lambda}(X) - \exp[-\lambda H_{\lambda+1}(X)] \\
 & \quad - \int_{\alpha}^{\beta} \mathcal{G}(y) (\lambda(1-\lambda)\alpha^{\lambda-2} - \lambda(1-\lambda)(2-\lambda)\alpha^{\lambda-3}(y-\alpha)) dy \\
 & \leq \lambda(1-\lambda)(2-\lambda)(3-\lambda) \left(\frac{\beta^{(\lambda-4)p+1} - \alpha^{(\lambda-4)p+1}}{(\lambda-4)p+1} \right)^{\frac{1}{p}} \\
 & \quad \times \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y,u) dy \right)^q du \right)^{\frac{1}{q}}, \tag{4.4}
 \end{aligned}$$

where

$$\mathcal{G}(y) = \sum_{i=1}^m p_i G(p_i, y) - G(E(X), y). \tag{4.5}$$

(ii) For $\lambda \in (1, \infty)$, we have

$$\begin{aligned} 0 &\leq \exp[-\lambda H_{\lambda+1}(X)] - (E(X))^\lambda \\ &\quad - \int_\alpha^\beta \mathcal{G}(y) (\lambda(\lambda - 1)\alpha^{\lambda-2} + \lambda(\lambda - 1)(\lambda - 2)\alpha^{\lambda-3}(y - \alpha)) dy \\ &\leq \lambda(\lambda - 1)|(\lambda - 2)(\lambda - 3)| \left(\frac{\beta^{(\lambda-4)p+1} - \alpha^{(\lambda-4)p+1}}{(\lambda - 4)p + 1} \right)^{\frac{1}{p}} \\ &\quad \times \left(\int_\alpha^\beta \left(\int_\alpha^\beta \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned}$$

where $\mathcal{G}(y)$ is defined by (4.5).

Proof (i) Substituting p_i in place of ξ_i in (4.2) and taking into account that

$$\sum_{i=1}^m p_i^\lambda = \exp[(1 - \lambda)H_\lambda(X)],$$

i.e.,

$$\sum_{i=1}^m p_i^{\lambda+1} = \exp[-\lambda H_{\lambda+1}(X)],$$

we get (4.4).

(ii) Similar to (i). □

Let \mathbf{u}, \mathbf{v} be two positive probability distributions. The following measures are well known in information theory:

- *Hellinger discrimination:*

$$h^2(\mathbf{u}, \mathbf{v}) = \frac{1}{2} \sum_{i=1}^m (\sqrt{u_i} - \sqrt{v_i})^2.$$

- χ^2 -divergence:

$$D_{\chi^2}(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^m \frac{(u_i - v_i)^2}{v_i}.$$

- *Triangular discrimination:*

$$\Delta(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^m \frac{(u_i - v_i)^2}{u_i + v_i}.$$

In the following results, we consider positive probability distributions $\mathbf{u}, \mathbf{v}, \mathbf{w}$ with the assumption of existence of a row stochastic matrix $\mathbf{A} \in \mathcal{M}_m(\mathbb{R})$ such that

$$\mathbf{w} = \mathbf{w}\mathbf{A} \quad \text{and} \quad \frac{\mathbf{v}}{\mathbf{w}} = \frac{\mathbf{u}}{\mathbf{w}}\mathbf{A}^T, \tag{4.6}$$

where $\frac{\mathbf{u}}{\mathbf{w}} = (\frac{u_1}{w_1}, \dots, \frac{u_m}{w_m})$ and $\frac{\mathbf{v}}{\mathbf{w}} = (\frac{v_1}{w_1}, \dots, \frac{v_m}{w_m})$.

Corollary 7 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be positive probability distributions such that $\frac{\mathbf{u}}{\mathbf{w}}, \frac{\mathbf{v}}{\mathbf{w}} \in [\alpha, \beta]^m$ and (4.6) is satisfied for some row stochastic matrix $\mathbf{A} \in \mathcal{M}_m(\mathbb{R})$. Then:*

(i)

$$\begin{aligned} 0 &\leq h^2(\mathbf{u}, \mathbf{w}) - h^2(\mathbf{v}, \mathbf{w}) - \int_{\alpha}^{\beta} \mathcal{G}(y) \left[\frac{1}{4} \alpha^{-\frac{3}{2}} - \frac{3}{8} \alpha^{-\frac{5}{2}} (y - \alpha) \right] dy \\ &\leq \frac{15}{16} \left[\frac{2}{2-7p} (\beta^{\frac{2-7p}{2}} - \alpha^{\frac{2-7p}{2}}) \right]^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{4.7}$$

where

$$\mathcal{G}(y) = \sum_{i=1}^m w_i G\left(\frac{u_i}{w_i}, y\right) - \sum_{i=1}^m w_i G\left(\frac{v_i}{w_i}, y\right). \tag{4.8}$$

(ii)

$$D_{\chi^2}(\mathbf{u}, \mathbf{w}) - D_{\chi^2}(\mathbf{v}, \mathbf{w}) = 2 \int_{\alpha}^{\beta} \mathcal{G}(u) du, \tag{4.9}$$

where $\mathcal{G}(y)$ is defined by (4.8).

(iii)

$$\begin{aligned} 0 &\leq \Delta(\mathbf{u}, \mathbf{w}) - \Delta(\mathbf{v}, \mathbf{w}) - \int_{\alpha}^{\beta} \mathcal{G}(y) [8(\alpha + 1)^{-3} - 24(\alpha + 1)^{-4}(y - \alpha)] dy \\ &\leq 96 \left[\frac{(\beta + 1)^{1-5p} - (\alpha + 1)^{1-5p}}{1 - 5p} \right]^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{4.10}$$

where $\mathcal{G}(y)$ is defined by (4.8).

Proof If we substitute x_i by $\frac{u_i}{w_i}$, y_i by $\frac{v_i}{w_i}$, a_i by w_i in (3.9) and

- (i) take $\phi(x) = \frac{1}{2}(\sqrt{x} - 1)^2$, we obtain (4.7).
- (ii) take $\phi(x) = (x - 1)^2$, we obtain (4.9).
- (iii) take $\phi(x) = \frac{(x-1)^2}{x+1}$, we obtain (4.10). □

5 Applications as the Zipf–Mandelbrot law

The Zipf–Mandelbrot law is a discrete probability distribution depending on three parameters $m \in \mathbb{N}, t \geq 0$, and $s > 0$ with probability mass function defined by

$$f(k, m, t, s) = \frac{1}{(k + t)^s H_{m,t,s}}, \quad k = 1, 2, \dots, m,$$

where

$$H_{m,t,s} = \sum_{i=1}^m \frac{1}{(i+t)^s}.$$

When $t = 0$, we get so-called Zipf’s law.

The Zipf–Mandelbrot, as well as Zipf’s, law has wide applications in many branches of science as well as linguistics [15], information sciences [16, 17], ecological field studies [18], etc. For more information, see also [15, 19].

We introduce the following definitions of Csiszár divergence for the Zipf–Mandelbrot law. For more information about Csiszár divergence, see [20, 21].

Definition 1 (Csiszár divergence for Z–M law) Let $m \in \mathbb{N}$ and $\phi : [\alpha, \beta] \rightarrow \mathbb{R}$ be a function.

(a) For $t \geq 0$ and $s, r_1, \dots, r_m > 0$, such that

$$\frac{1}{r_i(i+t)^s H_{m,t,s}} \in [\alpha, \beta], \quad i = 1, \dots, m,$$

we define

$$\hat{I}_\phi(m, t, s, \mathbf{r}) = \sum_{i=1}^m r_i \phi \left(\frac{1}{r_i(i+t)^s H_{m,t,s}} \right).$$

Specially, when $r_i = 1, i = 1, \dots, m$, we have

$$\hat{I}_\phi(m, t, s, \mathbf{1}) = \sum_{i=1}^m \phi \left(\frac{1}{(i+t)^s H_{m,t,s}} \right).$$

(b) For $t, \tilde{t} \geq 0$ and $s, \tilde{s} > 0$, such that

$$\frac{(i+\tilde{t})^{\tilde{s}} H_{m,\tilde{t},\tilde{s}}}{(i+t)^s H_{m,t,s}} \in [\alpha, \beta], \quad i = 1, \dots, m,$$

we define

$$\tilde{I}_\phi(m, t, \tilde{t}, s, \tilde{s}) = \sum_{i=1}^m \frac{1}{(i+\tilde{t})^{\tilde{s}} H_{m,\tilde{t},\tilde{s}}} \phi \left(\frac{(i+\tilde{t})^{\tilde{s}} H_{m,\tilde{t},\tilde{s}}}{(i+t)^s H_{m,t,s}} \right).$$

Corollary 8 Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let $m \in \mathbb{N}, t_1, t_2, t_3 \geq 0$ and $s_1, s_2, s_3 > 0$ be such that

$$\frac{(i+t_2)^{s_2} H_{m,t_2,s_2}}{(i+t_1)^{s_1} H_{m,t_1,s_1}}, \frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_2)^{s_2} H_{m,t_2,s_2}} \in [\alpha, \beta], \quad i = 1, \dots, m, \tag{5.1}$$

and

$$\begin{aligned} \frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_2)^{s_2} H_{m,t_2,s_2}} &= \sum_{j=1}^m \frac{(j+t_3)^{s_3} H_{m,t_3,s_3}}{(j+t_1)^{s_1} H_{m,t_1,s_1}} a_{ij}, \quad i = 1, \dots, m, \\ \frac{1}{(j+t_3)^{s_3} H_{m,t_3,s_3}} &= \sum_{i=1}^m \frac{1}{(i+t_3)^{s_3} H_{m,t_3,s_3}} a_{ij}, \quad j = 1, \dots, m, \end{aligned} \tag{5.2}$$

hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$. Then, for every 4-convex function $\phi : [\alpha, \beta] \rightarrow \mathbb{R}$, we have

$$\begin{aligned} 0 &\leq \tilde{I}_\phi(i, m, t_1, t_3, s_1, s_3) - \tilde{I}_\phi(i, m, t_2, t_3, s_2, s_3) - \int_\alpha^\beta \mathcal{G}(y) (\phi''(\alpha) + \phi'''(\alpha)(y - \alpha)) dy \\ &\leq \|\phi^{(4)}\|_p \left(\int_\alpha^\beta \left(\int_\alpha^\beta \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{5.3}$$

where

$$\begin{aligned} \mathcal{G}(y) &= \sum_{i=1}^m \frac{1}{(i+t_3)^{s_3} H_{m,t_3,s_3}} G\left(\frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_1)^{s_1} H_{m,t_1,s_1}}, y\right) \\ &\quad - \sum_{i=1}^m \frac{1}{(i+t_3)^{s_3} H_{m,t_3,s_3}} G\left(\frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_2)^{s_2} H_{m,t_2,s_2}}, y\right). \end{aligned} \tag{5.4}$$

Proof If we substitute x_i by $\frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_1)^{s_1} H_{m,t_1,s_1}}$, y_i by $\frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_2)^{s_2} H_{m,t_2,s_2}}$, and a_i by $\frac{1}{(i+t_3)^{s_3} H_{m,t_3,s_3}}$ in (3.9), we obtain the required result. \square

Corollary 9 Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let $m \in \mathbb{N}, t_1, t_2 \geq 0$ and $s_1, s_2, r_1, \dots, r_m > 0$ be such that

$$\frac{1}{r_i(i+t_1)^{s_1} H_{m,t_1,s_1}}, \frac{1}{r_i(i+t_2)^{s_2} H_{m,t_2,s_2}} \in [\alpha, \beta], \quad i = 1, \dots, m \tag{5.5}$$

and

$$\begin{aligned} \frac{1}{r_i(i+t_2)^{s_2} H_{m,t_2,s_2}} &= \sum_{j=1}^m \frac{1}{r_j(j+t_1)^{s_1} H_{m,t_1,s_1}} a_{ij}, \quad i = 1, \dots, m, \\ r_j &= \sum_{i=1}^m r_i a_{ij}, \quad j = 1, \dots, m, \end{aligned} \tag{5.6}$$

hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$. Then, for every 4-convex function $\phi : [\alpha, \beta] \rightarrow \mathbb{R}$, we have

$$\begin{aligned} 0 &\leq \hat{I}_\phi(m, t_1, s_1, \mathbf{r}) - \hat{I}_\phi(m, t_2, s_2, \mathbf{r}) - \int_\alpha^\beta \mathcal{G}(y) (\phi''(\alpha) + \phi'''(\alpha)(y - \alpha)) dy \\ &\leq \|\phi^{(4)}\|_p \left(\int_\alpha^\beta \left(\int_\alpha^\beta \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{5.7}$$

where

$$\mathcal{G}(y) = \sum_{i=1}^m r_i G\left(\frac{1}{r_i(i+t_1)^{s_1} H_{m,t_1,s_1}}, y\right) - \sum_{i=1}^m r_i G\left(\frac{1}{r_i(i+t_2)^{s_2} H_{m,t_2,s_2}}, y\right). \tag{5.8}$$

Proof If we substitute x_i by $\frac{1}{r_i(i+t_1)^{s_1} H_{m,t_1,s_1}}$, y_i by $\frac{1}{r_i(i+t_2)^{s_2} H_{m,t_2,s_2}}$, and a_i by $r_i > 0$ in (3.9), we obtain the required result. \square

Corollary 10 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G, G_{2,1}$ be defined by (2.1), (3.8), respectively. Let $m \in \mathbb{N}, t_1, t_2 \geq 0$, and $s_1, s_2 > 0$ be such that*

$$\frac{1}{(i+t_1)^{s_1} H_{m,t_1,s_1}}, \frac{1}{(i+t_2)^{s_2} H_{m,t_2,s_2}} \in [\alpha, \beta], \quad i = 1, \dots, m \tag{5.9}$$

and

$$\frac{1}{(i+t_2)^{s_2} H_{m,t_2,s_2}} = \sum_{j=1}^m \frac{1}{(j+t_1)^{s_1} H_{m,t_1,s_1}} a_{ij}, \quad i = 1, \dots, m, \tag{5.10}$$

hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$. Then, for every 4-convex function $\phi : [\alpha, \beta] \rightarrow \mathbb{R}$, we have

$$\begin{aligned} 0 &\leq \hat{I}_\phi(m, t_1, s_1, \mathbf{1}) - \hat{I}_\phi(m, t_2, s_2, \mathbf{1}) - \int_\alpha^\beta \mathcal{G}(y) (\phi''(\alpha) + \phi'''(\beta)(y - \alpha)) dy \\ &\leq \|\phi^{(4)}\|_p \left(\int_\alpha^\beta \left(\int_\alpha^\beta \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned}$$

where

$$\mathcal{G}(y) = \sum_{i=1}^m G\left(\frac{1}{(i+t_1)^{s_1} H_{m,t_1,s_1}}, y\right) - \sum_{i=1}^m G\left(\frac{1}{(i+t_2)^{s_2} H_{m,t_2,s_2}}, y\right). \tag{5.11}$$

Proof Substituting $r_i = 1, i = 1, \dots, m$, in (5.7), we get the required result. \square

Next we introduce definitions of Shannon’s entropy for the Zipf–Mandelbrot law.

Definition 2 (Shannon’s entropy for Z–M law) Let $m \in \mathbb{N}$.

(a) For $t \geq 0$ and $s, r_1, \dots, r_m > 0$, we define

$$\hat{H}(m, t, s, \mathbf{r}) = - \sum_{i=1}^m r_i \ln\left(\frac{1}{r_i(i+t)^s H_{m,t,s}}\right).$$

(b) For $t, \tilde{t} \geq 0$ and $s, \tilde{s} > 0$, we define

$$\tilde{H}(m, t, \tilde{t}, s, \tilde{s}) = - \sum_{i=1}^m \frac{1}{(i+\tilde{t})^{\tilde{s}} H_{m,\tilde{t},\tilde{s}}} \ln\left(\frac{(i+\tilde{t})^{\tilde{s}} H_{m,\tilde{t},\tilde{s}}}{(i+t)^s H_{m,t,s}}\right).$$

Corollary 11 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G_{2,1}$ be defined by (3.8) and $m \in \mathbb{N}$.*

- (i) *If $t_1, t_2, t_3 \geq 0$ and $s_1, s_2, s_3 > 0$ are such that (5.1) and (5.2) hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$, then*

$$0 \leq \tilde{H}(m, t_1, t_3, s_1, s_3) - \tilde{H}(m, t_2, t_3, s_2, s_3) - \int_{\alpha}^{\beta} \mathcal{G}(y) \frac{3\alpha - 2y}{\alpha^3} dy$$

$$\leq 6 \left(\frac{\beta^{1-4p} - \alpha^{1-4p}}{1 - 4p} \right)^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}},$$

where $\mathcal{G}(y)$ is defined by (5.4).

- (ii) *If $t_1, t_2 \geq 0$ and $s_1, s_2, r_1, \dots, r_m > 0$ are such that (5.5) and (5.6) hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$, then*

$$0 \leq \hat{H}(m, t_1, s_1, \mathbf{r}) - \hat{H}(m, t_2, s_2, \mathbf{r}) - \int_{\alpha}^{\beta} \mathcal{G}(y) \frac{3\alpha - 2y}{\alpha^3} dy$$

$$\leq 6 \left(\frac{\beta^{1-4p} - \alpha^{1-4p}}{1 - 4p} \right)^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \tag{5.12}$$

where $\mathcal{G}(y)$ is defined by (5.8).

- (iii) *If $t_1, t_2 \geq 0$ and $s_1, s_2 > 0$ are such that (5.9) and (5.10) hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$, then*

$$0 \leq \hat{H}(m, t_1, s_1, \mathbf{1}) - \hat{H}(m, t_2, s_2, \mathbf{1}) - \int_{\alpha}^{\beta} \mathcal{G}(y) \frac{3\alpha - 2y}{\alpha^3} dy$$

$$\leq 6 \left(\frac{\beta^{1-4p} - \alpha^{1-4p}}{1 - 4p} \right)^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}},$$

where $\mathcal{G}(y)$ is defined by (5.11).

Proof (i) Substituting x_i by $\frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_1)^{s_1} H_{m,t_1,s_1}}$, y_i by $\frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_2)^{s_2} H_{m,t_2,s_2}}$, and a_i by $\frac{1}{(i+t_3)^{s_3} H_{m,t_3,s_3}}$ and taking $\phi(x) = -\ln x$ in (3.9), we get the required result.

(ii) Substituting x_i by $\frac{1}{r_i(i+t_1)^{s_1} H_{m,t_1,s_1}}$, y_i by $\frac{1}{r_i(i+t_2)^{s_2} H_{m,t_2,s_2}}$, a_i by $r_i > 0$ in (3.9) and taking $\phi(x) = -\ln x$, we get the required result.

(iii) Substituting $r_i = 1$ in (5.12), we get the required result. □

At the end, we introduce the Kullback–Leibler divergence for the Zipf–Mandelbrot law. For more information about the Kullback–Leibler divergence, see [22, 23].

Definition 3 (The Kullback–Leibler divergence for Z–M) *Let $m \in \mathbb{N}$.*

- (a) *For $t \geq 0$ and $s, r_1, \dots, r_m > 0$, we define*

$$\hat{KL}(m, t, s, \mathbf{r}) = \sum_{i=1}^m \frac{1}{(i+t)^s H_{m,t,s}} \ln \left(\frac{1}{r_i(i+t)^s H_{m,t,s}} \right).$$

Specially, when $r_i = 1, i = 1, \dots, m$, we have

$$\hat{K}L(m, t, s, \mathbf{1}) = \sum_{i=1}^m \frac{1}{(i+t)^s H_{m,t,s}} \ln \left(\frac{1}{(i+t)^s H_{m,t,s}} \right).$$

(b) For $t, \tilde{t} \geq 0$ and $s, \tilde{s} > 0$, we define

$$\tilde{K}L(m, t, \tilde{t}, s, \tilde{s}) = \sum_{i=1}^m \frac{1}{(i+t)^s H_{m,t,s}} \ln \left(\frac{(i+\tilde{t})^{\tilde{s}} H_{m,\tilde{t},\tilde{s}}}{(i+t)^s H_{m,t,s}} \right).$$

Corollary 12 *Let p, q be a pair of conjugate exponents, i.e., $1 < p, q < \infty, 1/p + 1/q = 1$. Let $G_{2,1}$ be defined by (3.8) and $m \in \mathbb{N}$.*

(i) *If $t_1, t_2, t_3 \geq 0$ and $s_1, s_2, s_3 > 0$ are such that (5.1) and (5.2) hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$, then*

$$\begin{aligned} 0 &\leq \tilde{K}L(m, t_1, t_3, s_1, s_3) - \tilde{K}L(m, t_2, t_3, s_2, s_3) - \int_{\alpha}^{\beta} \mathcal{G}(y) \frac{2\alpha - y}{\alpha^2} dy \\ &\leq 2 \left(\frac{\beta^{1-3p} - \alpha^{1-3p}}{1 - 3p} \right)^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned}$$

where $\mathcal{G}(y)$ is defined by (5.4).

(ii) *If $t_1, t_2 \geq 0$ and $s_1, s_2, r_1, \dots, r_m > 0$ are such that (5.5) and (5.6) hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$, then*

$$\begin{aligned} 0 &\leq \hat{K}L(m, t_1, s_1, \mathbf{r}) - \hat{K}L(m, t_2, s_2, \mathbf{r}) - \int_{\alpha}^{\beta} \mathcal{G}(y) \frac{2\alpha - y}{\alpha^2} dy \\ &\leq 2 \left(\frac{\beta^{1-3p} - \alpha^{1-3p}}{1 - 3p} \right)^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned} \tag{5.13}$$

where $\mathcal{G}(y)$ is defined by (5.8).

(iii) *If $t_1, t_2 \geq 0$ and $s_1, s_2 > 0$ are such that (5.9) and (5.10) hold for some row stochastic matrix $A = (a_{ij}) \in \mathcal{M}_m(\mathbb{R})$, then*

$$\begin{aligned} 0 &\leq \hat{K}L(m, t_1, s_1, \mathbf{1}) - \hat{K}L(m, t_2, s_2, \mathbf{1}) - \int_{\alpha}^{\beta} \mathcal{G}(y) \frac{2\alpha - y}{\alpha^2} dy \\ &\leq 2 \left(\frac{\beta^{1-3p} - \alpha^{1-3p}}{1 - 3p} \right)^{\frac{1}{p}} \left(\int_{\alpha}^{\beta} \left(\int_{\alpha}^{\beta} \mathcal{G}(y) G_{2,1}(y, u) dy \right)^q du \right)^{\frac{1}{q}}, \end{aligned}$$

where $\mathcal{G}(y)$ is defined by (5.11).

Proof (i) Substituting x_i by $\frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_1)^{s_1} H_{m,t_1,s_1}}$, y_i by $\frac{(i+t_3)^{s_3} H_{m,t_3,s_3}}{(i+t_2)^{s_2} H_{m,t_2,s_2}}$, a_i by $\frac{1}{(i+t_3)^{s_3} H_{m,t_3,s_3}}$ in (3.9) and taking $\phi(x) = x \ln x$, we get the required result.

(ii) Substituting x_i by $\frac{1}{r_i (i+t_1)^{s_1} H_{m,t_1,s_1}}$, y_i by $\frac{1}{r_i (i+t_2)^{s_2} H_{m,t_2,s_2}}$, and a_i by r_i for each $i = 1, \dots, m$ in (3.9) and taking $\phi(x) = x \ln x$, we get the required result.

(iii) Substituting $r_i = 1, i = 1, \dots, m$, in (5.13), we get the required result. □

6 Conclusions

In this paper we have given generalized results for Sherman's inequality by considering the class of convex functions of higher order. We obtained an extended weighted majorization inequality as well as Jensen's inequality as special cases directly connected to information theory. We used the obtained results to derive new estimates for Shannon's and Rényi's entropy, information energy, and some well-known measures between probability distributions. Using the Zipf–Mandelbrot law, we introduced new functionals to derive some related results.

Acknowledgements

The publication was supported by the Ministry of Education and Science of the Russian Federation (the Agreement number No. 02.a03.21.0008). This publication is partially supported by the Royal Commission for Jubail and Yanbu, Kingdom of Saudi Arabia. The authors would like to thank an anonymous referee for their valuable remarks and suggestions that improved an earlier version of the manuscript.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors contributed equally. All authors read and approved the final manuscript.

Author details

¹Faculty of Civil Engineering, Architecture And Geodesy, University of Split, Split, Croatia. ²Department of General Studies, Jubail Industrial College, Jubail Industrial City, Kingdom of Saudi Arabia. ³Catholic University of Croatia, Zagreb, Croatia. ⁴Faculty of Textile Technology Zagreb, University of Zagreb, Zagreb, Croatia. ⁵RUDN University, Moscow, Russia.

Publisher's Note

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

Received: 3 November 2017 Accepted: 14 April 2018 Published online: 24 April 2018

References

1. Pečarić, J., Proschan, F., Tong, Y.L.: *Convex Functions, Partial Orderings and Statistical Applications*. Academic Press, New York (1992)
2. Hardy, G.H., Littlewood, J.E., Pólya, G.: *Inequalities*, 2nd edn. Cambridge University Press, Cambridge (1952)
3. Sherman, S.: On a theorem of Hardy, Littlewood, Pólya and Blackwell. *Proc. Natl. Acad. Sci. USA* **37**(1), 826–831 (1957)
4. Marshall, A.W., Olkin, I., Arnold, B.C.: *Inequalities: Theory of Majorization and Its Application*, 2nd edn. Springer Series in Statistics. Springer, New York (2011)
5. Agarwal, R.P., Wong, P.J.Y.: *Error Inequalities in Polynomial Interpolation and Their Applications*. Kluwer Academic, Dordrecht (1993)
6. Shannon, C.E., Weaver, W.: *The Mathematical Theory of Communication*, Urbana. University of Illinois Press, Champaign (1949)
7. Budimir, I., Dragomir, S.S., Pečarić, J.: Further reverse results for Jensen's discrete inequality and applications in information theory. *J. Inequal. Pure Appl. Math.* **2**(1), 5 (2001)
8. Dragomir, S.S.: A converse of the Jensen inequality for convex mappings of several variables and applications. *Acta Math. Vietnam.* **29**(1), 77–88 (2004)
9. Dragomir, S.S., Dragomir, N.M., Praneş, K.: An inequality for logarithms and applications in information theory. *Comput. Math. Appl.* **38**, 11–17 (1999)
10. Matić, M., Pearce, C.E.M., Pečarić, J.: Improvements of some bounds on entropy measures in information theory. *Math. Inequal. Appl.* **1**, 295–304 (1998)
11. Matić, M., Pearce, C.E.M., Pečarić, J.: On an inequality for the entropy of a probability distribution. *Acta Math. Hung.* **85**, 345–349 (1999)
12. Matić, M., Pearce, C.E.M., Pečarić, J.: Refinements of some bounds in information theory. *ANZIAM J.* **42**, 387–398 (2001)
13. Matić, M., Pearce, C.E.M., Pečarić, J.: Shannon's and related inequalities in information theory. In: *Survey on Classical Inequalities*, pp. 127–164. Kluwer Academic, Norwell (2000)
14. Rényi, A.: On measures of entropy and information. In: *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, vol. 1, pp. 547–561. University of California Press, Berkeley (1961)
15. Montemurro, M.A.: Beyond the Zipf–Mandelbrot law in quantitative linguistics. arXiv:cond-mat/0104066v2 (2001)
16. Egghe, L., Rousseau, R.: *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science*. Elsevier, New York (1990)
17. Silagadze, Z.K.: Citations and the Zipf–Mandelbrot law. *Complex Syst.* **11**, 487–499 (1997)
18. Mouillot, D., Lepretre, A.: Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity. *Environ. Monit. Assess.* **63**(2), 279–295 (2000)
19. Mandelbrot, B.: Information theory and psycholinguistics: a theory of words frequencies. In: Lazafeld, P., Henry, N. (eds.) *Reading in Mathematical Social Science*. MIT Press, Cambridge (1966)

20. Csiszár, I.: Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hung.* **2**, 299–318 (1967)
21. Csiszár, I.: Information measure: a critical survey. In: *Trans. 7th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Processes and 8th European Meeting of Statist.*, B, pp. 73–86. Academia, Prague (1978)
22. Kullback, S.: *Information Theory and Statistics*. Wiley, New York (1959)
23. Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
