Mathematical
Inequalities
& Applications
Volume 22, Number 4 (2019), 1405–1419 doi:10.7153/mia-2019-22-98

CONVERSE TO THE SHERMAN INEQUALITY WITH APPLICATIONS

ANA BARBIR, SLAVICA IVELIĆ BRADANOVIĆ, ĐILDA PEČARIĆ AND JOSIP PEČARIĆ

(*Communicated by J. Jakšetić*)

Abstract. In this paper we proved a converse to Sherman's inequality. Using the concept of f-divergence we obtained some inequalities for the well-known entropies. We also introduced a new entropy by applying the Zipf-Mandelbrot law and derived some related inequalities.

1. Introduction and preliminaries

Throughout $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the sets of nonnegative and positive numbers, i.e. $\mathbb{R}_+ = [0,\infty)$ and $\mathbb{R}_{++} = (0,\infty)$, respectively.

Let $f : [\alpha,\beta] \to \mathbb{R}$ be a convex function on $[\alpha,\beta] \subset \mathbb{R}$. If $\boldsymbol{x} = (x_1,\ldots,x_n)$ is any n-tuple in $[\alpha,\beta]^n$ and $\boldsymbol{a} = (a_1,\ldots,a_n) \in \mathbb{R}^n_+$ such that $\sum_{i=1}^n a_i = 1$, then the well known Jensen inequality

$$f\left(\sum_{i=1}^n a_i x_i\right) \leqslant \sum_{i=1}^n a_i f(x_i) \tag{1.1}$$

holds (see for example [18]).

Closely connected to Jensen's inequality (1.1) is the Lah-Ribarič inequality

$$\sum_{i=1}^n a_i f(x_i) \leqslant \frac{\beta-\bar{x}}{\beta-\alpha} f(\alpha) + \frac{\bar{x}-\alpha}{\beta-\alpha} f(\beta), \tag{1.2}$$

which holds for every function $f : [\alpha,\beta] \to \mathbb{R}$ convex on $[\alpha,\beta] \subset \mathbb{R}$, where $\mathbf{x} = (x_1,\ldots,x_n) \in [\alpha,\beta]^n$, $\mathbf{a} = (a_1,\ldots,a_n) \in \mathbb{R}^n_+$ with $\sum_{i=1}^n a_i = 1$ and $\bar{x} = \sum_{i=1}^n a_i x_i$ (see [16]).

Sherman [21] obtained generalization of Jensen's inequality (1.1) in the form

$$\sum_{j=1}^m b_j f(y_j) \leqslant \sum_{i=1}^n a_i f(x_i), \tag{1.3}$$

Mathematics subject classification (2010): 94A17, 26D15, 15B51.
Keywords and phrases: Sherman inequality, Majorization inequality, Jensen inequality, f-divergence, Zipf-Mandelbrot entropy, convex function.

Paper MIA-22-98

which holds for every function $f : [\alpha,\beta] \to \mathbb{R}$ convex on $[\alpha,\beta] \subset \mathbb{R}$, where $\mathbf{x} = (x_1,\ldots,x_n) \in [\alpha,\beta]^n$, $\mathbf{y} = (y_1,...,y_m) \in [\alpha,\beta]^m$, $\mathbf{a} = (a_1,...,a_n) \in \mathbb{R}^n_+$, and $\mathbf{b} = (b_1,...,b_m) \in \mathbb{R}^m_+$ are such that

$$\mathbf{y} = \mathbf{x}S \text{ and } \mathbf{a} = \mathbf{b}S^{\mathsf{T}} \tag{1.4}$$

for some column stochastic matrix $S = (s_{ij}) \in \mathscr{M}_{nm}(\mathbb{R})$, i.e. matrix whose entries are greater or equal to zero with the sum of the entries in each column is equal to 1. Here S^{T} denotes a transpose matrix of S.

Recently, some generalization of Sherman's inequality (1.3) are obtained (see [1, 2, 7–11, 17]).

Note that (1.4) can be written as

$$\begin{aligned}
\mathbf{y} = \mathbf{x}S, &\quad \left(y_j = \sum_{i=1}^{n} x_i s_{ij},\;\; j = 1,...,m\right), \\
\mathbf{a} = \mathbf{b}S^{\mathsf{T}}, &\quad \left(a_i = \sum_{j=1}^{m} b_j s_{ij},\;\; i = 1,...,n\right).
\end{aligned} \tag{1.5}$$

It is obvious that Sherman's inequality (1.3) reduces to Jensen's inequality (1.1) by choosing $m = 1$ and setting $\mathbf{b} = [1]$.

Csiszár [4] introduced the concept of f-divergence functional

$$C_f(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right) \tag{1.6}$$

for a convex function $f : \mathbb{R}_{++} \to \mathbb{R}$ and $\mathbf{p} = (p_1,...,p_n) \in \mathbb{R}^n_{++}$, $\mathbf{q} = (q_1,...,q_n) \in \mathbb{R}^n_{++}$.

It is possible to use non-negative n-tuples $\mathbf{p}$ and $\mathbf{q}$ in the f-divergence functional, by defining

$$f(0) = \lim_{t\to 0+} f(t), \quad 0f\left(\frac{0}{0}\right) = 0, \quad 0f\left(\frac{c}{0}\right) = \lim_{\varepsilon\to 0+} f\left(\frac{c}{\varepsilon}\right) = c\lim_{t\to\infty}\frac{f(t)}{t}, \quad c > 0.$$

We will limit our consideration to positive cases of $\mathbf{p}$ and $\mathbf{q}$.

The generalized Csiszár f-divergence for a convex function $f : \mathbb{R}_{++} \to \mathbb{R}$ is defined by

$$C_f(\mathbf{p},\mathbf{q};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right), \tag{1.7}$$

where $\mathbf{p} = (p_1,...,p_n) \in \mathbb{R}_{++}$, $\mathbf{q} = (q_1,...,q_n) \in \mathbb{R}_{++}$, with weights $\mathbf{r} = (r_1,...,r_n) \in \mathbb{R}_+$. It is obvious $C_f(\mathbf{p},\mathbf{q};\mathbf{e}) = C_f(\mathbf{p},\mathbf{q})$ for $\mathbf{e} = (1,...,1) \in \mathbb{R}^n$.

The classical inequality for f-divergence functional, known as the Csiszár-Körner inequality [5], has the form

$$\sum_{i=1}^{n} p_i f\left(\frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} p_i}\right) \leqslant C_f(\mathbf{p},\mathbf{q}) \tag{1.8}$$

and holds for every function $f:\mathbb{R}_{++}\to\mathbb{R}$ convex on $\mathbb{R}_{++}$. Specially, if f is normalized, i.e. $f(1)=0$ and $\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i$, then

$$0\leqslant C_f(\mathbf{p},\mathbf{q}). \tag{1.9}$$

In particular, if $\mathbf{p}$ and $\mathbf{q}$ are two positive probability distribution, i.e. $\mathbf{p}=(p_1,\ldots,p_n)\in\mathbb{R}_{++}^{n}$ and $\mathbf{q}=(q_1,\ldots,q_n)\in\mathbb{R}_{++}^{n}$ with $\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i=1$, then the inequality (1.9) holds for every convex and normalized function $f:\mathbb{R}_{++}\to\mathbb{R}$. These results are easy consequences of Jensen's inequality (1.1).

In this paper, as main result we present a converse to Sherman's inequality (1.3). Using the concept of f-divergence we also obtain a converse to the Csiszár-Körner inequality (1.8). As easy consequences we derive some inequalities for the well-known divergences. As applications, we introduce a new entropy by applying the Zipf--Mandelbrot law and give some related inequalities including the Zipf-Mandelbrot entropy.

2. Main results

First we present a converse to Sherman's inequality (1.3).

THEOREM 1. *Let $f:[\alpha,\beta]\to\mathbb{R}$ be a convex function on $[\alpha,\beta]\subset\mathbb{R}$. Let $\mathbf{x}=(x_1,\ldots,x_n)\in[\alpha,\beta]^n$, $\mathbf{y}=(y_1,\ldots,y_m)\in[\alpha,\beta]^m$, $\mathbf{a}=(a_1,\ldots,a_n)\in\mathbb{R}_{+}^{n}$ and $\mathbf{b}=(b_1,\ldots,b_m)\in\mathbb{R}_{+}^{m}$ be such that (1.4) holds for some column stochastic matrix $S=(s_{ij})\in\mathscr{M}_{nm}(\mathbb{R})$, then*

$$\sum_{j=1}^{m}b_jf(y_j)\leqslant\sum_{i=1}^{n}a_if(x_i)\leqslant\sum_{j=1}^{m}b_j\frac{f(\alpha)(\beta-y_j)+f(\beta)(y_j-\alpha)}{\beta-\alpha}. \tag{2.1}$$

Proof. Under the assumptions, Sherman's inequality (1.3) holds. Further, from (1.2), setting $p_i=s_{ij}$, for $i=1,\ldots,n$, we have

$$\begin{aligned}\sum_{j=1}^{m}b_jf(y_j)&\leqslant\sum_{i=1}^{n}a_if(x_i)=\sum_{i=1}^{n}\left(\sum_{j=1}^{m}b_js_{ij}\right)f(x_i)=\sum_{j=1}^{m}b_j\left(\sum_{i=1}^{n}s_{ij}f(x_i)\right)\\&\leqslant\sum_{j=1}^{m}b_j\left(\frac{\beta-\sum_{i=1}^{n}x_is_{ij}}{\beta-\alpha}f(\alpha)+\frac{\sum_{i=1}^{n}x_is_{ij}-\alpha}{\beta-\alpha}f(\beta)\right),\end{aligned}$$

what we need to prove. □

In sequel, we use notation $\langle\cdot,\cdot\rangle$ for the standard inner product in $\mathbb{R}^n$. We also denote with $\mathscr{M}_{nm}(\mathbb{R}_+)$ the space of $n\times m$ matrices with nonnegative entries.

By applying Theorem 1 we compare two generalized Csiszár f-divergences.

THEOREM 2. *Let $f:[\alpha,\beta]\to\mathbb{R}$ be a convex function on $[\alpha,\beta]\subset\mathbb{R}_{++}$. Let $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\dots,n$. Further, let $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$, $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$, $\mathbf{c}\in\mathbb{R}^n_{+}$ and $\mathbf{d}\in\mathbb{R}^m_{+}$ be such that*

$$\tilde{\mathbf{p}}=\mathbf{p}R,\quad \tilde{\mathbf{q}}=\mathbf{q}R \quad \text{and} \quad \mathbf{c}=\mathbf{d}R^{\mathsf{T}} \tag{2.2}$$

for some matrix $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$, then

$$C_f(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d})\leqslant C_f(\mathbf{p},\mathbf{q};\mathbf{c})\leqslant\sum_{j=1}^m d_j\langle\mathbf{p},\mathbf{r}_j\rangle\frac{f(\alpha)\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+f(\beta)\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{2.3}$$

Proof. According to (1.7) the inequality (2.3) can be written in the form

$$\begin{aligned}\sum_{j=1}^m d_j\tilde{p}_j f\left(\frac{\tilde{q}_j}{\tilde{p}_j}\right)&\leqslant\sum_{i=1}^n c_ip_if\left(\frac{q_i}{p_i}\right)\\&\leqslant\sum_{j=1}^m d_j\langle\mathbf{p},\mathbf{r}_j\rangle\frac{f(\alpha)\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+f(\beta)\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}.\end{aligned} \tag{2.4}$$

We denote $\mathbf{r}_j=(r_{1j},\dots,r_{nj})$, $r_{ij}\geqslant 0$ for $i=1,\dots,n$, $j=1,\dots,m$. From (2.2) it follows that $\tilde{p}_j=\langle\mathbf{p},\mathbf{r}_j\rangle=\sum_{i=1}^n p_ir_{ij}$ and $\tilde{q}_j=\langle\mathbf{q},\mathbf{r}_j\rangle=\sum_{i=1}^n q_ir_{ij}$ for $j=1,\dots,m$. Moreover, $c_i=\sum_{j=1}^m d_jr_{ij}$ for $i=1,\dots,n$ (see (2.2)) and after multiplying with p_i and taking $a_i=c_ip_i$, $b_j=d_j\langle\mathbf{p},\mathbf{r}_j\rangle$ we get

$$a_i=\sum_{j=1}^m b_j\frac{p_ir_{ij}}{\langle\mathbf{p},\mathbf{r}_j\rangle} \tag{2.5}$$

for $i=1,\dots,n$, $j=1,\dots,m$. The following equality holds

$$\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}=\frac{p_1r_{1j}}{\sum_{i=1}^n p_irij}\frac{q_1}{p_1}+\dots+\frac{p_nr_{nj}}{\sum_{i=1}^n p_irij}\frac{q_n}{p_n}$$

for $j=1,\dots,m$. Hence, the following identity is valid

$$\left[\frac{\langle\mathbf{q},\mathbf{r}_1\rangle}{\langle\mathbf{p},\mathbf{r}_1\rangle},\dots,\frac{\langle\mathbf{q},\mathbf{r}_m\rangle}{\langle\mathbf{p},\mathbf{r}_m\rangle}\right]=\left[\frac{q_1}{p_1},\dots,\frac{q_n}{p_n}\right]\begin{bmatrix}\frac{p_1r_{11}}{\langle\mathbf{p},\mathbf{r}_1\rangle}&\cdots&\frac{p_1r_{1m}}{\langle\mathbf{p},\mathbf{r}_m\rangle}\\\vdots&\ddots&\vdots\\\frac{p_nr_{n1}}{\langle\mathbf{p},\mathbf{r}_1\rangle}&\cdots&\frac{p_nr_{nm}}{\langle\mathbf{p},\mathbf{r}_m\rangle}\end{bmatrix}. \tag{2.6}$$

The $n\times m$ matrix $S=(s_{ij})$, $s_{ij}=\frac{p_ir_{ij}}{\langle\mathbf{p},\mathbf{r}_j\rangle}$ is column stochastic and with $\mathbf{x}=(x_1,\dots,x_n)$, $\mathbf{y}=(y_1,\dots,y_n)$, $x_i=\frac{q_i}{p_i}$ and $y_j=\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}$, $i=1,\dots,n$, $j=1,\dots,m$, satisfies condition $\mathbf{y}=\mathbf{x}S$ (see (2.6)). Since $\mathbf{a}=\mathbf{b}S^{\mathsf{T}}$ (see (2.5)) is satisfied for $\mathbf{a}=(a_1,\dots,a_n)$ and $\mathbf{b}=$

$(b_1,...,b_m)$, we can apply Theorem 1 and obtain

$$\sum_{j=1}^{m} b_j f\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) = \sum_{j=1}^{m} d_j\langle \mathbf{p},\mathbf{r}_j\rangle f\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) \leqslant \sum_{i=1}^{n} c_i p_i f\left(\frac{q_i}{p_i}\right)$$
$$\leqslant \frac{\sum_{j=1}^{m} d_j\langle \mathbf{p},\mathbf{r}_j\rangle\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)}{\beta-\alpha} f(\alpha) + \frac{\sum_{j=1}^{m} d_j\langle \mathbf{p},\mathbf{r}_j\rangle\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha} f(\beta),$$

which is equivalent to (2.3). □

COROLLARY 1. *Let $f:[\alpha,\beta]\to\mathbb{R}$ be a convex function on $[\alpha,\beta]\subset\mathbb{R}_{++}$. Let $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,...,n$. Further, let $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$ and $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$ be such that*

$$\tilde{\mathbf{p}} = \mathbf{p}R \quad \textit{and} \quad \tilde{\mathbf{q}} = \mathbf{q}R \tag{2.7}$$

for some matrix $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$ and $\mathbf{R}=(R_1,...,R_n)$, where $R_i=\sum_{j=1}^m r_{ij}$, $i=1,...,n$ is the i-th row sum of R, then

$$C_f(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant C_f(\mathbf{p},\mathbf{q};\mathbf{R}) \leqslant \sum_{j=1}^{m}\langle \mathbf{p},\mathbf{r}_j\rangle \frac{f(\alpha)\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+f(\beta)\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{2.8}$$

In particular, if the matrix R is row stochastic, then

$$C_f(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant C_f(\mathbf{p},\mathbf{q}) \leqslant \sum_{j=1}^{m}\langle \mathbf{p},\mathbf{r}_j\rangle \frac{f(\alpha)\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+f(\beta)\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{2.9}$$

Proof. By taking $\mathbf{d}=(d_1,...,d_m)=(1,...,1)$ in Theorem 2, we calculate $c_i=\sum_{j=1}^m r_{ij}=R_i$ for $i=1,...,n$. Therefore inequality (2.3) becomes (2.8).
If additionally the matrix R is row stochastic, then $\mathbf{R}=(1,...,1)\in\mathbb{R}^n$ and (2.8) reduces to (2.9). □

As a special case of the previous result we obtain a converse to the Csiszár-Körner inequality (1.8).

COROLLARY 2. *Let $f:[\alpha,\beta]\to\mathbb{R}$ be a convex function on $[\alpha,\beta]\subset\mathbb{R}_{++}$. Let $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, $\mathbf{r}\in\mathbb{R}^n_{+}$ be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,...,n$, with $P_n=\sum_{i=1}^n p_i$ and $Q_n=\sum_{i=1}^n q_i$, then*

$$\langle \mathbf{p},\mathbf{r}\rangle f\left(\frac{\langle \mathbf{q},\mathbf{r}\rangle}{\langle \mathbf{p},\mathbf{r}\rangle}\right) \leqslant C_f(\mathbf{p},\mathbf{q};\mathbf{r}) \leqslant \langle \mathbf{p},\mathbf{r}\rangle \frac{f(\alpha)\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}\rangle}{\langle \mathbf{p},\mathbf{r}\rangle}\right)+f(\beta)\left(\frac{\langle \mathbf{q},\mathbf{r}\rangle}{\langle \mathbf{p},\mathbf{r}\rangle}-\alpha\right)}{\beta-\alpha}. \tag{2.10}$$

In particular, if $\mathbf{r}=\mathbf{e}$, then

$$P_n f\left(\frac{Q_n}{P_n}\right) \leqslant C_f(\mathbf{p},\mathbf{q}) \leqslant P_n \frac{f(\alpha)\left(\beta-\frac{Q_n}{P_n}\right)+f(\beta)\left(\frac{Q_n}{P_n}-\alpha\right)}{\beta-\alpha}. \tag{2.11}$$

Proof. Taking $m=1$ in Corollary 1 and $\mathbf{r}_1=(r_1,\dots,r_n)$, we obtain $R_i=r_i$ for $i=1,\dots,n$, and (2.8) becomes (2.10). Further, for $\mathbf{r}=\mathbf{e}=(1,\dots,1)$, the inequality (2.10) reduces to (2.11). $\square$

3. Application to divergences

In the examples below we obtain, for suitable choices of the kernel f, some of the best known distance functions used in mathematical statistics, information theory and other scientic fields (see [3, 6, 12–15, 19, 20]).

For $f(t)=-\ln t,\ t>0$, the Csiszáre f-divergence is

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

known as *the Kullback-Liebler divergence*.

We also introduce *the weighted Kullback-Liebler divergence* defined by

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

with $r_i\geqslant 0,\ i=1,\dots,n$. Obviously, for $\mathbf{e}=(1,\dots,1)$, it follows $KL(\mathbf{p},\mathbf{q};\mathbf{e})=KL(\mathbf{p},\mathbf{q})$.

The Shannon entropy is defined by

$$H(\mathbf{p})=-\sum_{i=1}^{n}p_i\ln p_i, \tag{3.1}$$

where $\mathbf{p}\in\mathbb{R}^n_{++}$. Note that the Shannon entropy we can get as a special case from the Csiszáre f-divergence choosing the convex mapping $f(t)=\ln\frac{1}{t}=-\ln t, t>0$, i.e.

$$C_f(\mathbf{p},\mathbf{e})=-\sum_{i=1}^{n}p_i\ln\left(\frac{1}{p_i}\right)=\sum_{i=1}^{n}p_i\ln p_i=-H(\mathbf{p}).$$

We also consider *the weighted Shannon entropy* defined by

$$H(\mathbf{p};\mathbf{r})=-\sum_{i=1}^{n}r_ip_i\ln p_i, \tag{3.2}$$

with weights r_i, $i=1,\dots,n$. Obviously, for $\mathbf{r}=\mathbf{e}=(1,\dots,1)$, it follows $H(\mathbf{p};\mathbf{e})=H(\mathbf{p})$.

COROLLARY 3. *Let $[\alpha,\beta]\subset\mathbb{R}_{++}$, $\mathbf{p}\in\mathbb{R}^n_{++}$ and $\mathbf{q}\in\mathbb{R}^n_{++}$ be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\dots,n$. Let $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$, $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$, $\mathbf{c}\in\mathbb{R}^n_{+}$ and $\mathbf{d}\in\mathbb{R}^m_{+}$ be such that (2.2) holds for some matrix $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$, then*

$$KL(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d})\leqslant KL(\mathbf{p},\mathbf{q};\mathbf{c})\leqslant\sum_{j=1}^{m}d_j\langle\mathbf{p},\mathbf{r}_j\rangle\frac{\ln\left(\frac{1}{\alpha}\right)\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+\ln\left(\frac{1}{\beta}\right)\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.3}$$

Proof. If we take in Theorem 2 function f to be $f(t) = \ln\left(\frac{1}{t}\right)$, which is convex on $[\alpha, \beta]$, then (3.3) follows from (2.3). $\square$

COROLLARY 4. *Let* $[\alpha,\beta] \subset \mathbb{R}_{++}$, $\mathbf{p} \in \mathbb{R}^n_{++}$, $\mathbf{q} \in \mathbb{R}^n_{++}$, *be such that* $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1,\dots,n$, *and* $\tilde{\mathbf{p}} \in \mathbb{R}^m_{++}$, $\tilde{\mathbf{q}} \in \mathbb{R}^m_{++}$ *be such that (2.7) holds for some matrix* $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$. *Further, let* $\mathbf{R} = (R_1,\dots,R_n)$, *where* $R_i = \sum_{j=1}^m r_{ij}$, $i = 1,\dots,n$ *is the i-th row sum of* R, *then*

$$KL(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant KL(\mathbf{p},\mathbf{q};\mathbf{R}) \leqslant \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\ln\left(\frac{1}{\alpha}\right)\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + \ln\left(\frac{1}{\beta}\right)\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}. \quad (3.4)$$

In particular, if the matrix R *is row stochastic, then*

$$KL(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant KL(\mathbf{p},\mathbf{q}) \leqslant \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\ln\left(\frac{1}{\alpha}\right)\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + \ln\left(\frac{1}{\beta}\right)\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}. \quad (3.5)$$

Proof. By taking $\mathbf{d} = (d_1,\dots,d_m) = (1,\dots,1)$ in Theorem 2 we obtain $c_i = \sum_{j=1}^m r_{ij} = R_i$ for $i = 1,\dots,n$. Therefore inequality (2.3) becomes (3.4).
If additionally R is row stochastic, then $\mathbf{R} = (1,\dots,1) \in \mathbb{R}^n$ and (2.8) becomes (3.5). $\square$

COROLLARY 5. *Let* $[\alpha,\beta] \subset \mathbb{R}_{++}$, $\mathbf{p} \in [\alpha,\beta]^n$, $\tilde{\mathbf{p}} \in [\alpha,\beta]^m$, $\mathbf{c} \in \mathbb{R}^n_+$ *and* $\mathbf{d} \in \mathbb{R}^m_+$ *be such that*

$$\tilde{\mathbf{p}} = \mathbf{p}R \text{ and } \mathbf{c} = \mathbf{d}R^{\mathsf{T}}$$

for some column stochastic matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$, *then*

$$H(\tilde{\mathbf{p}};\mathbf{d}) \geqslant H(\mathbf{p};\mathbf{c}) \geqslant \sum_{j=1}^m d_j \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\ln(\alpha)\left(\beta - \frac{1}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + \ln(\beta)\left(\frac{1}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}. \quad (3.6)$$

Proof. We take in Theorem 2 a function f to be $f(t) = \ln\frac{1}{t}$ which is convex on $[\alpha,\beta]$ and $\mathbf{q} = \mathbf{e} = (1,\dots,1) \in \mathbb{R}^m$. Then, since R is column stochastic, we also have $\tilde{\mathbf{q}} = (\langle \mathbf{q},\mathbf{r}_1\rangle,\dots,\langle \mathbf{q},\mathbf{r}_m\rangle) = (\langle \mathbf{e},\mathbf{r}_1\rangle,\dots,\langle \mathbf{e},\mathbf{r}_m\rangle) = (1,\dots,1)$. Then (3.6) follows from (2.3). $\square$

COROLLARY 6. *Let* $[\alpha,\beta] \subset \mathbb{R}_{++}$, $\mathbf{p} \in [\alpha,\beta]^n$ *and* $\tilde{\mathbf{p}} \in [\alpha,\beta]^m$ *be such that*

$$\tilde{\mathbf{p}} = \mathbf{p}R \quad (3.7)$$

for some column stochastic matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$ *and* $\mathbf{R} = (R_1,\dots,R_n)$, *where* $R_i = \sum_{j=1}^m r_{ij}$, $i = 1,\dots,n$ *is the i-th row sum of* R, *then*

$$H(\tilde{\mathbf{p}}) \geqslant H(\mathbf{p};\mathbf{R}) \geqslant \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\ln(\alpha)\left(\beta - \frac{1}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + \ln(\beta)\left(\frac{1}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}. \quad (3.8)$$

In particular, if the matrix R is double stochastic, then

$$H(\tilde{\mathbf{p}}) \geqslant H(\mathbf{p}) \geqslant \sum_{j=1}^{m} \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{\ln(\alpha)\left(\beta - \frac{1}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + \ln(\beta)\left(\frac{1}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{\beta - \alpha}. \tag{3.9}$$

Proof. By taking $\mathbf{d} = (d_1, ..., d_m) = (1, ..., 1)$ in Theorem 2 we obtain $c_i = \sum_{j=1}^{m} r_{ij} = R_i$ for $i = 1, ..., n$. Therefore inequality (2.3) becomes (3.8).
If additionally the matrix R is row stochastic, then $\mathbf{R} = (1, ..., 1) \in \mathbb{R}^n$ and (2.8) becomes (3.9). □

Consider now *the Hellinger distance*

$$h(\mathbf{p}, \mathbf{q}) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2}, \tag{3.10}$$

where $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n_{++}$. This distance is metric and is often used in its squared form

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

We also define *the weighted Hellinger distance,* with weights $\mathbf{r} = (r_1, ..., r_n) \in \mathbb{R}_+$, in squared form

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

We know that Hellinger disctance is actually the Csiszáre f-divergence for the convex mapping $f(t) = \frac{1}{2}\left(1 - \sqrt{t}\right)^2$.

COROLLARY 7. *Let $[\alpha, \beta] \subset \mathbb{R}_{++}$, $\mathbf{p} \in \mathbb{R}^n_{++}$, $\mathbf{q} \in \mathbb{R}^n_{++}$, be such that $\frac{q_i}{p_i} \in [\alpha, \beta]$, $i = 1, ..., n$, and $\tilde{\mathbf{p}} \in \mathbb{R}^m_{++}$, $\tilde{\mathbf{q}} \in \mathbb{R}^m_{++}$, $\mathbf{c} \in \mathbb{R}^n_+$, $\mathbf{d} \in \mathbb{R}^m_+$ be such that (2.2) holds for some matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$, then*

$$\begin{aligned} h^2(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}; \mathbf{d}) &\leqslant h^2(\mathbf{p}, \mathbf{q}; \mathbf{c}) \\ &\leqslant \sum_{j=1}^{m} d_j \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{(1 - \sqrt{\alpha})^2 \left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + (1 - \sqrt{\beta})^2 \left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{2(\beta - \alpha)}. \end{aligned} \tag{3.11}$$

Proof. If we take in Theorem 2 function f to be $f(t) = \frac{1}{2}\left(1 - \sqrt{t}\right)^2$ which is convex on $[\alpha, \beta]$, equation (3.11) follows from (2.3). □

COROLLARY 8. *Let $[\alpha, \beta] \subset \mathbb{R}_{++}$, $\mathbf{p} \in \mathbb{R}^n_{++}$, $\mathbf{q} \in \mathbb{R}^n_{++}$, be such that $\frac{q_i}{p_i} \in [\alpha, \beta]$, $i = 1, ..., n$, and $\tilde{\mathbf{p}} \in \mathbb{R}^m_{++}$, $\tilde{\mathbf{q}} \in \mathbb{R}^m_{++}$ be such that (2.7) holds for some matrix $R = (r_{ij}) \in$*

$\mathscr{M}_{nm}(\mathbb{R}_+)$ and $\mathbf{R}=(R_1,\dots,R_n)$, where $R_i=\sum_{j=1}^m r_{ij}$, $i=1,\dots,n$ is the i-th row sum of R, then

$$h^2(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant h^2(\mathbf{p},\mathbf{q};\mathbf{R}) \leqslant \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{(1-\sqrt{\alpha})^2\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+(1-\sqrt{\beta})^2\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{2(\beta-\alpha)}. \tag{3.12}$$

In particular, if the matrix R is row stochastic, then

$$h^2(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant h^2(\mathbf{p},\mathbf{q}) \leqslant \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{(1-\sqrt{\alpha})^2\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+(1-\sqrt{\beta})^2\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{2(\beta-\alpha)}. \tag{3.13}$$

Proof. By taking $\mathbf{d}=(d_1,\dots,d_m)=(1,\dots,1)$ in Theorem 2 we obtain $c_i=\sum_{j=1}^m r_{ij}$ $=R_i$ for $i=1,\dots,n$. Therefore inequality (2.3) becomes (3.4).
If additionally the matrix R is row stochastic, then $\mathbf{R}=(1,\dots,1)\in\mathbb{R}^n$ and (2.8) becomes (3.5). □

For the convex function $f(t)=-\sqrt{t}$ and $\mathbf{p},\mathbf{q}\in\mathbb{R}^n_{++}$, we get

$$C_f(\mathbf{p},\mathbf{q})=\sum_{i=1}^n p_i\left(-\sqrt{\frac{q_i}{p_i}}\right)=-\sum_{i=1}^n\sqrt{p_iq_i}=-B(\mathbf{p},\mathbf{q}),$$

known as the Bhattacharyya distance.

COROLLARY 9. *Let $[\alpha,\beta]\subset\mathbb{R}_{++}$, $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\dots,n$. Let $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$, $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$, $\mathbf{c}\in\mathbb{R}^n_+$ and $\mathbf{d}\in\mathbb{R}^m_+$ be such that (2.2) holds for some matrix $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$, then*

$$B(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d}) \geqslant B(\mathbf{p},\mathbf{q};\mathbf{c}) \geqslant \sum_{j=1}^m d_j\langle \mathbf{p},\mathbf{r}_j\rangle \frac{\sqrt{\alpha}\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+\sqrt{\beta}\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.14}$$

Proof. If we take in Theorem 2 function f to be $f(t)=-\sqrt{t}$, which is convex on $[\alpha,\beta]$, equation (3.14) follows from (2.3). □

COROLLARY 10. *Let $[\alpha,\beta]\subset\mathbb{R}_{++}$, $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\dots,n$. Let $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$, $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$ be such that (2.7) holds for some matrix $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$ and $\mathbf{R}=(R_1,\dots,R_n)$, where $R_i=\sum_{j=1}^m r_{ij}$, $i=1,\dots,n$ is the i-th row sum of R, then*

$$B(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \geqslant B(\mathbf{p},\mathbf{q};\mathbf{R}) \geqslant \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\sqrt{\alpha}\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+\sqrt{\beta}\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.15}$$

In particular, if the matrix R is row stochastic, then

$$B(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \geqslant B(\mathbf{p},\mathbf{q}) \geqslant \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\sqrt{\alpha}\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+\sqrt{\beta}\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.16}$$

Proof. By taking $\mathbf{d} = (d_1, \dots, d_m) = (1, \dots, 1)$ in Theorem 2 we obtain $c_i = \sum_{j=1}^{m} r_{ij} = R_i$ for $i = 1, \dots, n$. Therefore inequality (2.3) becomes (3.15).
If additionally R is row stochastic, then $\mathbf{R} = (1, \dots, 1) \in \mathbb{R}^n$ and (2.8) becomes (3.16). □

For suitable choices of a convex function f we define divergences as follows:
For $f(t) = (1-t)^2, t > 0$, we obtain χ^2*-divergence*

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

For $f(t) = |1-t|, t > 0$, we obtain *the total variation distance*

$$C_f(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} p_i \left|1 - \frac{q_i}{p_i}\right| = \sum_{i=1}^{n} |p_i - q_i| = V(\mathbf{p},\mathbf{q}).$$

For $f(t) = \frac{(1-t)^2}{t+1}, t > 0$, we obtain *the triangular discrimination*

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

We also introduce their weighted versions, with weights $r_i \geqslant 0, i = 1, ..., n$:

$$\begin{aligned}
\chi^2(\mathbf{p},\mathbf{q};\mathbf{r}) &= \sum_{i=1}^{n} r_i \frac{(p_i - q_i)^2}{p_i}, \\
V(\mathbf{p},\mathbf{q};\mathbf{r}) &= \sum_{i=1}^{n} r_i |p_i - q_i|, \\
\Delta(\mathbf{p},\mathbf{q};\mathbf{r}) &= \sum_{i=1}^{n} r_i \frac{(p_i - q_i)^2}{p_i + q_i}.
\end{aligned}$$

COROLLARY 11. *Let* $[\alpha,\beta] \subset \mathbb{R}_{++}$, $\mathbf{p} \in \mathbb{R}^n_{++}$, $\mathbf{q} \in \mathbb{R}^n_{++}$, *be such that* $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1, ..., n$. *Let* $\tilde{\mathbf{p}} \in \mathbb{R}^m_{++}$, $\tilde{\mathbf{q}} \in \mathbb{R}^m_{++}$, $\mathbf{c} \in \mathbb{R}^n_{+}$ *and* $\mathbf{d} \in \mathbb{R}^m_{+}$ *be such that (2.2) holds for some matrix* $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$, *then*

$$\begin{aligned}
\chi^2(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d}) &\leqslant \chi^2(\mathbf{p},\mathbf{q};\mathbf{c}) \\
&\leqslant \sum_{j=1}^{m} d_j \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{(1-\alpha)^2 \left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + (1-\beta)^2 \left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{\beta - \alpha}. \qquad (3.17)
\end{aligned}$$

Proof. If we take in Theorem 2 function f to be $f(t) = (1-t)^2$ which is convex on $[\alpha,\beta]$, equation (3.17) follows from (2.3). □

COROLLARY 12. *Let* $[\alpha,\beta] \subset \mathbb{R}_{++}$, $\mathbf{p} \in \mathbb{R}^n_{++}$, $\mathbf{q} \in \mathbb{R}^n_{++}$, *be such that* $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1, ..., n$. *Let* $\tilde{\mathbf{p}} \in \mathbb{R}^m_{++}$, $\tilde{\mathbf{q}} \in \mathbb{R}^m_{++}$ *be such that (2.7) holds for some matrix* $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$ *and* $\mathbf{R} = (R_1, ..., R_n)$, *where* $R_i = \sum_{j=1}^{m} r_{ij}$, $i = 1, ..., n$ *is the* *i-th row sum of* R*, then*

$$\chi^2(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant \chi^2(\mathbf{p},\mathbf{q};\mathbf{R}) \leqslant \sum_{j=1}^{m}\langle \mathbf{p},\mathbf{r}_j\rangle \frac{(1-\alpha)^2\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+(1-\beta)^2\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.18}$$

In particular, if the matrix R *is row stochastic, then*

$$\chi^2(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant \chi^2(\mathbf{p},\mathbf{q}) \leqslant \sum_{j=1}^{m}\langle \mathbf{p},\mathbf{r}_j\rangle \frac{(1-\alpha)^2\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+(1-\beta)^2\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.19}$$

Proof. By taking $\mathbf{d}=(d_1,\ldots,d_m)=(1,\ldots,1)$ in Theorem 2 we obtain $c_i=\sum_{j=1}^{m} r_{ij}=R_i$ for $i=1,\ldots,n$. Therefore inequality (2.3) becomes (3.18).
If additionally R is row stochastic, then $\mathbf{R}=(1,\ldots,1)\in\mathbb{R}^n$ and (2.8) becomes (3.19). □

COROLLARY 13. *Let* $[\alpha,\beta]\subset\mathbb{R}_{++}$, $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, *be such that* $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\ldots,n$. *Further, let* $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$, $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$, $\mathbf{c}\in\mathbb{R}^n_{+}$ *and* $\mathbf{d}\in\mathbb{R}^m_{+}$ *be such that (2.2) holds for some matrix* $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$*, then*

$$\begin{aligned} V(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d}) &\leqslant V(\mathbf{p},\mathbf{q};\mathbf{c})\\ &\leqslant \sum_{j=1}^{m} d_j\langle \mathbf{p},\mathbf{r}_j\rangle \frac{|1-\alpha|\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+|1-\beta|\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \end{aligned} \tag{3.20}$$

Proof. If we take in Theorem 2 function f to be $f(t)=|1-t|$ which is convex on $[\alpha,\beta]$, equation (3.20) follows from (2.3). □

COROLLARY 14. *Let* $[\alpha,\beta]\subset\mathbb{R}_{++}$, $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, *be such that* $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\ldots,n$. *Let* $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$ *and* $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$ *be such that (2.7) holds for some matrix* $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$. *Further, let* $\mathbf{R}=(R_1,\ldots,R_n)$*, where* $R_i=\sum_{j=1}^{m} r_{ij}$, $i=1,\ldots,n$ *is the i-th row sum of* R*, then*

$$V(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant V(\mathbf{p},\mathbf{q};\mathbf{R}) \leqslant \sum_{j=1}^{m}\langle \mathbf{p},\mathbf{r}_j\rangle \frac{|1-\alpha|\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+|1-\beta|\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.21}$$

In particular, if the matrix R *is row stochastic, then*

$$V(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leqslant V(\mathbf{p},\mathbf{q}) \leqslant \sum_{j=1}^{m}\langle \mathbf{p},\mathbf{r}_j\rangle \frac{|1-\alpha|\left(\beta-\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right)+|1-\beta|\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.22}$$

Proof. By taking $\mathbf{d}=(d_1,\ldots,d_m)=(1,\ldots,1)$ in Theorem 2 we obtain $c_i=\sum_{j=1}^{m} r_{ij}=R_i$ for $i=1,\ldots,n$. Therefore inequality (2.3) becomes (3.21).
If additionally R is row stochastic, then $\mathbf{R}=(1,\ldots,1)\in\mathbb{R}^n$ and (2.8) becomes (3.22). □

COROLLARY 15. *Let* $[\alpha,\beta]\subset\mathbb{R}_{++}$, $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, *be such that* $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\dots,n$. *Further, let* $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$, $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$, $\mathbf{c}\in\mathbb{R}^n_{+}$ *and* $\mathbf{d}\in\mathbb{R}^m_{+}$ *be such that (2.2) holds for some matrix* $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$*, then*

$$\Delta(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d})\leqslant\Delta(\mathbf{p},\mathbf{q};\mathbf{c})\leqslant\sum_{j=1}^m d_j\langle\mathbf{p},\mathbf{r}_j\rangle\frac{\frac{(1-\alpha)^2}{\alpha+1}\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+\frac{(1-\beta)^2}{\beta+1}\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.23}$$

Proof. If we take in Theorem 2 function f to be $f(t)=\frac{(1-t)^2}{t+1}$ which is convex on $[\alpha,\beta]$, equation (3.23) follows from (2.3). □

COROLLARY 16. *Let* $[\alpha,\beta]\subset\mathbb{R}_{++}$, $\mathbf{p}\in\mathbb{R}^n_{++}$, $\mathbf{q}\in\mathbb{R}^n_{++}$, *be such that* $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\dots,n$. *Let* $\tilde{\mathbf{p}}\in\mathbb{R}^m_{++}$, $\tilde{\mathbf{q}}\in\mathbb{R}^m_{++}$ *be such that (2.7) holds for some matrix* $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$ *and* $\mathbf{R}=(R_1,\dots,R_n)$*, where* $R_i=\sum_{j=1}^m r_{ij}$, $i=1,\dots,n$ *is the i-th row sum of* R*, then*

$$\Delta(\tilde{\mathbf{p}},\tilde{\mathbf{q}})\leqslant\Delta(\mathbf{p},\mathbf{q};\mathbf{R})\leqslant\sum_{j=1}^m\langle\mathbf{p},\mathbf{r}_j\rangle\frac{\frac{(1-\alpha)^2}{\alpha+1}\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+\frac{(1-\beta)^2}{\beta+1}\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.24}$$

In particular, if the matrix R *is row stochastic, then*

$$\Delta(\tilde{\mathbf{p}},\tilde{\mathbf{q}})\leqslant\Delta(\mathbf{p},\mathbf{q})\leqslant\sum_{j=1}^m\langle\mathbf{p},\mathbf{r}_j\rangle\frac{\frac{(1-\alpha)^2}{\alpha+1}\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+\frac{(1-\beta)^2}{\beta+1}\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{3.25}$$

Proof. By taking $\mathbf{d}=(d_1,\dots,d_m)=(1,\dots,1)$ in Theorem 2 we obtain $c_i=\sum_{j=1}^m r_{ij}=R_i$ for $i=1,\dots,n$. Therefore inequality (2.3) becomes (3.24).
If additionally R is row stochastic, then $\mathbf{R}=(1,\dots,1)\in\mathbb{R}^n$ and (2.8) becomes (3.25). □

4. Inequalities 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

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

where

$$H_{n,q,s}=\sum_{i=1}^n\frac{1}{(i+q)^s}. \tag{4.1}$$

Using the given Zipf-Mandelbrot law we define new entropy by

$$Z(H,q,s)=\frac{s}{H_{n,q,s}}\sum_{k=1}^n\frac{\ln(k+q)}{(k+q)^s}+\ln H_{n,q,s}. \tag{4.2}$$

We also consider the weighted Zipf-Mandelbrot entropy defined by

$$Z(H,q,s,\mathbf{R}) = \frac{s}{H_{n,q,s,\mathbf{R}}} \sum_{k=1}^{n} R_k \frac{\ln(k+q)}{(k+q)^s} + \ln H_{n,q,s,\mathbf{R}}, \tag{4.3}$$

with nonnegative weights $R_i,\ i = 1,...,n$ and

$$H_{n,q,s,\mathbf{R}} = \sum_{i=1}^{n} \frac{R_i}{(i+q)^s}. \tag{4.4}$$

Specially, when r_{ij} are entries of some matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$, we use notation

$$H_{n,q,s,\mathbf{r}_j} = \sum_{i=1}^{n} \frac{r_{ij}}{(i+q)^s}. \tag{4.5}$$

THEOREM 3. *Let $n \in \mathbb{N}$, $q \geqslant 0$ and $s > 0$. Let $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$ be some column stochastic matrix, $\mathbf{R} = (R_1,\ldots,R_n)$, where $R_i = \sum_{j=1}^{m} r_{ij},\ i = 1,\ldots,n$ is the i-th row sum of R, then*

$$\begin{aligned}\sum_{j=1}^{m} \frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s,\mathbf{R}}} \ln\left(\frac{H_{n,q,s,\mathbf{R}}}{H_{n,q,s,\mathbf{r}_j}}\right) &\geqslant Z(H,q,s,\mathbf{R}) \\ &\geqslant \sum_{j=1}^{m} \frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s,\mathbf{R}}} \frac{\ln(\alpha)\left(\beta - \frac{H_{n,q,s,\mathbf{R}}}{H_{n,q,s,\mathbf{r}_j}}\right) + \ln(\beta)\left(\frac{H_{n,q,s,\mathbf{R}}}{H_{n,q,s,\mathbf{r}_j}} - \alpha\right)}{\beta - \alpha},\end{aligned} \tag{4.6}$$

provided that all terms are well defined.
In particular, if the matrix R is double stochastic, then

$$\begin{aligned}\sum_{j=1}^{m} \frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s}} \ln\left(\frac{H_{n,q,s}}{H_{n,q,s,\mathbf{r}_j}}\right) &\geqslant Z(H,q,s) \\ &\geqslant \sum_{j=1}^{m} \frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s}} \frac{\ln(\alpha)\left(\beta - \frac{H_{n,q,s}}{H_{n,q,s,\mathbf{r}_j}}\right) + \ln(\beta)\left(\frac{H_{n,q,s}}{H_{n,q,s,\mathbf{r}_j}} - \alpha\right)}{\beta - \alpha}.\end{aligned} \tag{4.7}$$

Proof. Since $H_{n,q,s,\mathbf{R}} = \sum_{i=1}^{n} \frac{R_i}{(i+q)^s}$, it is obvious that

$$\sum_{i=1}^{n} \frac{R_i}{(i+q)^s H_{n,q,s,\mathbf{R}}} = H_{n,q,s,\mathbf{R}} \cdot \frac{1}{H_{n,q,s,\mathbf{R}}} = 1.$$

If we substitute p_i with $\frac{1}{(i+q)^s H_{n,q,s,\mathbf{R}}}$, $i=1,2,...,n$, then

$$\begin{aligned}
H(\mathbf{p};\mathbf{R}) &= -\sum_{i=1}^{n} R_i p_i \ln p_i = -\sum_{i=1}^{n} \frac{R_i}{(i+q)^s H_{n,q,s,\mathbf{R}}} \ln \frac{1}{(i+q)^s H_{n,q,s,\mathbf{R}}} \\
&= \sum_{i=1}^{n} \frac{R_i}{(i+q)^s H_{n,q,s,\mathbf{R}}} \ln \left((i+q)^s H_{n,q,s,\mathbf{R}}\right) \\
&= \sum_{i=1}^{n} \frac{R_i \ln(i+q)^s}{(i+q)^s H_{n,q,s,\mathbf{R}}} + \sum_{i=1}^{n} \frac{R_i \ln H_{n,q,s,\mathbf{R}}}{(i+q)^s H_{n,q,s,\mathbf{R}}} \\
&= \frac{s}{H_{n,q,s,\mathbf{R}}} \sum_{i=1}^{n} \frac{R_i \ln(i+q)}{(i+q)^s} + \frac{\ln H_{n,q,s,\mathbf{R}}}{H_{n,q,s,\mathbf{R}}} \sum_{i=1}^{n} \frac{R_i}{(i+q)^s} \\
&= \frac{s}{H_{n,q,s,\mathbf{R}}} \sum_{i=1}^{n} \frac{R_i \ln(i+q)}{(i+q)^s} + \ln H_{n,q,s,\mathbf{R}} = Z(H,q,s,\mathbf{R}).
\end{aligned}$$

From $\tilde{\mathbf{p}} = \mathbf{p}R$, it follows

$$\tilde{p}_j = \langle \mathbf{p}, \mathbf{r}_j \rangle = \sum_{i=1}^{n} p_i r_{ij} = \sum_{i=1}^{n} \frac{r_{ij}}{(i+q)^s H_{n,q,s,\mathbf{R}}} = \frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s,\mathbf{R}}},$$

so we have

$$H(\tilde{\mathbf{p}}) = -\sum_{j=1}^{m} \tilde{p}_j \ln \tilde{p}_j = -\sum_{j=1}^{m} \frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s,\mathbf{R}}} \ln \left(\frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s,\mathbf{R}}} \right) = \sum_{j=1}^{m} \frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s,\mathbf{R}}} \ln \left(\frac{H_{n,q,s,\mathbf{R}}}{H_{n,q,s,\mathbf{r}_j}} \right).$$

Now applying (3.8) we get the required result.
Specially, if R is also row stochastic, then $\mathbf{R} = (1,\ldots,1) \in \mathbb{R}^n$. Further, we have $H_{n,q,s,\mathbf{R}} = H_{n,q,s}$ and $Z(H,q,s,\mathbf{R}) = Z(H,q,s)$, so the inequality (4.6) reduces to (4.7). □

Acknowledgement. The publication was supported by the Ministry of Education and Science of the Russian Federation (the Agreement number No. 02.a03.21.0008.). 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, J. PEČARIĆ, *Generalizations of Sherman's inequality by Hermite's interpolating polynomial*, Math. Inequal. Appl. **19** (4) (2016) 1181–1192.

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

[3] I. BURBEA AND C. R. RAO, *On the convexity of some divergence measures based on entropy functions*, IEEE Transactions on Information Theory, 28 (1982), 489–495.

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

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

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

[7] 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.

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

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

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

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

[12] J. H. JUSTICE, *Maximum Entropy and Bayssian Methods in Applied Statistics*, Cambridge University Press, Cambridge, 1986.

[13] J. N. KAPUR, *On the roles of maximum entropy and minimum discrimination information principles in Statistics*, Technical Address of the 38th Annual Conference of the Indian Society of Agricultural Statistics, 1984, 1–44.

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

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

[16] P. LAH AND M. RIBARIČ, *Converse of Jensen's inequality for convex functions*, Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat. Fiz. 412–460 (1973), 201–205.

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

[18] J. E. PEČARIĆ, F. PROSHAN AND Y. L. TONG, *Convex Functions, Partial Orderings and Statistical Applications*, Academic Press, Inc. (1992).

[19] A. RENYI, *On measures of entropy and information*, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 547–561, 1961.

[20] C. E. SHANNON, *A mathematical theory of communication*, Bell System Technical Journal 27 (1948) 379–423.

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

(Received November 1, 2018)

Ana Barbir
Faculty of Civil Engineering, Architecture And Geodesy
University of Split
Matice Hrvatske 15, 21000 Split, Croatia
e-mail: ana.barbir@gradst.hr

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

Josip Pečarić
RUDN University
Moscow, Russia
e-mail: pecaric@element.hr