MONOGRAPHS IN INEQUALITIES 15

Inequalities and Zipf-Mandelbrot Law

Selected topics in information theory

Edited by: Đilda Pečarić and Josip Pečarić

Inequalities and Zipf-Mandelbrot Law

Selected topics in information theory

Edited by:

Đilda Pečarić
Catholic University of Croatia
Ilica 242, 10 000 Zagreb, Croatia

Josip Pečarić
Zagreb, Croatia

Zagreb, 2019

MONOGRAPHS IN INEQUALITIES 15

Inequalities and Zipf-Mandelbrot Law

Selected topics in information theory

Edited by: Đilda Pečarić and Josip Pečarić

1st edition

Printed in Croatia.

A CIP catalogue record for this book is available from the National and University Library in Zagreb under 001022990.

ISBN 978-953-197-670-1

Preface

Inequalities and Zipf-Mandelbrot law is devoted to recent advances in variety of inequalities in Information theory especially for Zipf law and Zipf-Mandelbrot law. Since, Zipf-Mandelbrot law has been applied in various scientific disciplines and different kind of natural or social phenomena: from text mining, information retrieval, animal communication, to gene expression and many others, we hope that this book will be useful to different scientific communities.

Subjects covered in this volume include: Zipf-Mandelbrot law and hybrid Zipf-Mandelbrot law, Properties and its generalizations, Zipf-Mandelbrot entropy, Approximating f-divergence via Hermite interpolation polynomials, Bounds for Inequalities for Entropy of Zipf-Mandelbrot law, Inequalities for Shanon and Zipf-Mandelbrot entropies by using Jensen type inequalities, Combinatorial improvements of Zipf-Mandelbrot laws wia interpolations, Cyclic improvements of Inequalitires for Entropy of Zipf-Mandelbrot law, Inequalitires of the Jensen and Edmunson-Lah-Ribarič type for Zipf-Mandelbrot law, Sherman's inequality with applications in information theory, Jensen-type inequalities for generalized f-divergence and Zipf-Mandelbrot law as well as some related results for Shanon and Zipf-Mandelbrot entropies.

We wish to express our appreciation to the distinguished mathematicians who contributed to this volume and all the researchers who contributed to this specific field of research. Finally, it is our pleasure to acknowledge the fine cooperation and assistance provided by the staff of "Element",

Đilda Pečarić
Josip Pečarić

Editors

Contents

Chapter 1

Zipf-Mandelbrot law, properties and its generalizations

Julije Jakšetić, Đilda Pečarić and Josip Pečarić

Abstract. Despite a wide spread applications of Zipf-Mandelbrot law, there is quite small amount of results concerning analytical properties on distribution law. On the first stage, we examine some monotonicity properties of the law, we derive the whole variety of its lower and upper estimations. We then further refine our results using some well-known inequalities such as Hölder and Lyapunov inequality.

On the second stage we consider the case when total mass of Zipf-Mandelbrot law is spread all over positive integer, and then we come to Hurwitz $\zeta-$function. As we show, it is very natural first to examine properties of Hurwitz $\zeta-$function to derive properties of Zipf-Mandelbrot law. Using some well-known inequalities such as Chebyshev's and Lyapunov's inequality we are able to deduce a whole variety of theoretical characterizations that include, among others, log-convexity, log-subadditivity, exponential convexity.

On the third stage, we generalize Zipf-Mandelbrot law using maximization of Shannon entropy, as we get hybrid Zipf-Mandelbrot law. It is interesting that examination of its densities provides some new insights of Lerch's transcendent.

1.1 Some classical inequalities and Zipf-Mandelbrot law

1.1.1 Introduction

For $N \in \mathbb{N},\ q \geq 0,\ s > 0,\ k \in \{1,2,\ldots,N\}$, Zipf-Mandelbrot probability mass function is defined with

$$f(k,N,q,s) = \frac{1/(k+q)^s}{H_{N,q,s}}, \tag{1.1}$$

where

$$H_{N,q,s} = \sum_{i=1}^{N} \frac{1}{(i+q)^s}, \tag{1.2}$$

$N \in \mathbb{N},\ q \geq 0,\ s > 0,\ k \in \{1,2,\ldots,N\}$ (see [6]).

Proposition 1.1 *For* $s > t > 0$

$$(Nf(k,N,q,s))^{1/s} \leq (Nf(k,N,q,t))^{1/t}. \tag{1.3}$$

Proof. In [7] it is proved, after $\frac{1}{Nf(k,N,q,s)}$ is interpreted as power mean depending on s, that $s \mapsto Nf(k,N,q,s)$ is a decreasing function. □

Denote $m = \frac{k+q}{N+q}$, $M = \frac{k+q}{1+q}$ and observe $m = \min\{x_i :\ i = 1,\ldots,N\}$, $M = \max\{x_i :\ i = 1,\ldots,N\}$.
Further, for $s,\ t > 0$ let

$$\mu = \frac{M^s - m^s}{M^t - m^t}$$

and

$$B_{t,s} = \left(\frac{\mu t}{s}\right)^{\frac{1}{t}} \left\{ \frac{m^s M^t - m^t M^s}{(1-s/t)(M^t - m^t)} \right\}^{\frac{1}{s}-\frac{1}{t}}. \tag{1.4}$$

Theorem 1.1 *For probability mass function* (1.39) *we have following inequalities, for* $0 < t < s$

a)

$$\frac{N^{\frac{s}{t}-1}}{B_{t,s}^s} (f(k,N,q,t))^{s/t} \leq f(k,N,q,s) \leq N^{\frac{s}{t}-1} (f(k,N,q,t))^{s/t}, \tag{1.5}$$

b)

$$\frac{M^t - m^t}{f(k,N,q,s)} - \frac{M^s - m^s}{f(k,N,q,t)} \leq N\left(M^t m^s - M^s m^t\right). \tag{1.6}$$

Proof.
a) It follows, for $0 < t < s$,

$$(Nf(k,N,q,s))^{1/s} \le (Nf(k,N,q,t))^{1/t},$$

hence

$$f(k,N,q,s) \le N^{\frac{s}{t}-1}\,(f(k,N,q,t))^{s/t}.$$

Now we prove left hand side inequality. First, observe here that $m = \min\{x_i : i = 1,\ldots,N\}$, $M = \max\{x_i : i = 1,\ldots,N\}$.
Using *Beesack inequality (see [3], p.* 334; *[15]*, p. 110)

$$M_N^{[s]}(x_{\overline{1,N}}) \le B_{t,s} M_N^{[t]}(x_{\overline{1,N}}),\ 0 < t < s, \tag{1.7}$$

where

$$B_{t,s} = \left(\frac{\mu t}{s}\right)^{\frac{1}{t}} \left\{ \frac{m^s M^t - m^t M^s}{(1-s/t)\,(M^t - m^t)} \right\}^{\frac{1}{s}-\frac{1}{t}}.$$

It follows

$$f(k,N,q,s) \ge \frac{N^{\frac{s}{t}-1}}{B_{t,s}^s}\,(f(k,N,q,t))^{s/t}.$$

b) From *Goldman inequality* (see [15], p. 109.), $0 < t < s$,

$$\left(M^t - m^t\right)\{M_N^{[s]}(x_{\overline{1,N}})\}^s - (M^s - m^s)\,\{M_N^{[t]}(x_{\overline{1,N}})\}^t \le M^t m^s - M^s m^t.$$

Hence, for $0 < t < s$,

$$\frac{M^t - m^t}{f(k,N,q,s)} - \frac{M^s - m^s}{f(k,N,q,t)} \le N\left(M^t m^s - M^s m^t\right).$$

□

Remark 1.1 Another type of a lower bound for $f(k,N,q,s)$ can be derived from another Beesack inequality (see [3], p. 336; [15], p. 111):

$$M_N^{[s]}(x_{\overline{1,N}}) \le C_{t,s} + M_N^{[t]}(x_{\overline{1,N}}),$$

where

$$C_{t,s} = \left\{ \frac{m^s M^t}{M^t - m^t} + \frac{s-t}{t}\left(\frac{\mu t}{s}\right)^{\frac{s}{s-t}} \right\}^{\frac{1}{s}},$$

concluding

$$f(k,N,q,s) \ge \frac{1}{N} \cdot \frac{1}{\left(C_{t,s} + [Nf(k,N,q,t)]^{-\frac{1}{t}}\right)^s}.$$

1.1.2 Zipf law estimations

If we take $q = 0$ in probability mass function (1.39) we get Zipf law with probability mass function

$$f(k,N,s) = \frac{1}{k^s H_{N,s}} \tag{1.8}$$

where

$$H_{N,s} = \sum_{i=1}^{N} \frac{1}{i^s}. \tag{1.9}$$

For $s = 1$ $H_N = H_{N,1}$ we get $N-$th harmonic number.

1° (case $t = 1$)

Using Proposition 1.2 for $q = 0,\ t = 1$ and $s > 1$ we have

$$(Nf(k,N,s))^{\frac{1}{s}} \leq Nf(k,N,1)$$

i.e.

$$f(k,N,s) \leq \frac{N^{s-1}}{k^s H_N^s}. \tag{1.10}$$

We can derive further bounds using well-known inequalities for harmonic numbers. Using Schlömlich-Lemonnier inequalities (see [14], p. 118)

$$\ln(N+1) < H_N < 1 + \ln(N+1) \tag{1.11}$$

and (1.10) we get

$$f(k,N,s) < N^{s-1} k^{-s} \ln^{-s}(N+1).$$

Also, using (see [14], p. 120)

$$r(1-(N+1)^{-1/r}) < H_n < r(N^{1/r} - 1) + 1 \tag{1.12}$$

we have

$$f(k,N,s) < N^{s-1}(rk(1-(N+1)^{-1/r}))^{-s}.$$

Similarly, we have a list of inequalities with Euler constant $\gamma = \lim_{N\to}(H_N - \ln N)$ (see [14], p. 120):

$$\gamma + \ln N + \tfrac{1}{2N} - \tfrac{1}{8N^2} < H_N < \gamma + \ln N + \tfrac{1}{2N} \tag{1.13}$$

$$\gamma + \ln N + \tfrac{1}{2(N+1)} < H_N < \gamma + \ln N + \tfrac{1}{2(N-1)} \tag{1.14}$$

$$\gamma + \ln(N+1/2) + \tfrac{1}{24(N+1)^2} < H_N < \gamma + \ln(N+1/2) + \tfrac{1}{24N^2} \tag{1.15}$$

$$\begin{aligned} &\gamma + \ln(N+1/2) + \tfrac{1}{24(N+1/2)^2} - \tfrac{7}{960N^4} < H_N \\ &< \gamma + \ln(N+1/2) + \tfrac{1}{24(N+1/2)^2} - \tfrac{7}{960(N+1)^4}. \end{aligned} \tag{1.16}$$

Now, using (1.10) and left-hand side inequalities in (1.13)-(1.16) we get

$$f(k,N,s) < k^{-s} N^{s-1} \left(\gamma + \ln N + \tfrac{1}{2N} - \tfrac{1}{8N^2}\right)^{-s}$$

$$f(k,N,s) < k^{-s}N^{s-1}\left(\gamma + \ln N + \tfrac{1}{2(N+1)}\right)^{-s}$$
$$f(k,N,s) < k^{-s}N^{s-1}\left(\gamma + \ln(N+1/2) + \tfrac{1}{24(N+1)^2}\right)^{-s}$$
$$f(k,N,s) < k^{-s}N^{s-1}\left(\gamma + \ln(N+1/2) + \tfrac{1}{24(N+1/2)^2} - \tfrac{7}{960N^4}\right)^{-s}$$

Similarly, using Proposition 1.2 for $q = 0,\ t = 1$ and $0 < s < 1$ we have

$$(Nf(k,N,s))^{\frac{1}{s}} \geq Nf(k,N,1)$$

i.e.

$$f(k,N,s) \geq \frac{N^{s-1}}{k^s H_N^s}. \tag{1.17}$$

and then using (1.13)-(1.16) we will get lower bounds

$$f(k,N,s) > k^{-s}N^{s-1}\left(\gamma + \ln N + \tfrac{1}{2N}\right)^{-s}$$
$$f(k,N,s) > k^{-s}N^{s-1}\left(\gamma + \ln N + \tfrac{1}{2(N-1)}\right)^{-s}$$
$$f(k,N,s) > k^{-s}N^{s-1}\left(\gamma + \ln(N+1/2) + \tfrac{1}{24N^2}\right)^{-s}$$
$$f(k,N,s) > k^{-s}N^{s-1}\left(\gamma + \ln(N+1/2) + \tfrac{1}{24(N+1/2)^2} - \tfrac{7}{960(N+1)^4}\right)^{-s}$$

2° (case $t = 2$)
Using Proposition 1.2 for $q = 0,\ t = 2$ and $s > 2$ we have

$$(Nf(k,N,s))^{\frac{1}{s}} \leq Nf(k,N,1) = \left(Nk^{-2}H_{N,2}\right)^{\frac{1}{2}}$$

i.e.

$$f(k,N,s) \leq N^{\frac{s}{2}-1}k^{-s}H_{N,2}^{-\frac{s}{2}}. \tag{1.18}$$

Appling Proposition 1.2 for $q = 0,\ t = 2$ and $0 < s < 2$ we get reversed inequality

$$f(k,N,s) \geq N^{\frac{s}{2}-1}k^{-s}H_{N,2}^{-\frac{s}{2}}. \tag{1.19}$$

Now we use the next estimations for $H_{N,2}$ (see [14] p. 121–122; [16])

$$\frac{\pi^2}{6} - \frac{N+1/2}{N^2+N+d} < H_{N,2} < \frac{\pi^2}{6} - \frac{N+1/2}{N^2+N+1/3},\ d = 0.324555 \tag{1.20}$$

and (see [14] p. 122)

$$H_{N,2} \geq \tfrac{8}{5} - \tfrac{1}{N+\frac{2}{3}},\ N \geq 1 \tag{1.21}$$
$$H_{N,2} \geq \tfrac{13}{8} - \tfrac{1}{N+\frac{3}{5}},\ N \geq 1 \tag{1.22}$$
$$H_{N,2} \geq \tfrac{13}{8} - \tfrac{1}{N+\frac{2}{3}},\ N \geq 2 \tag{1.23}$$

$$H_{N,2} \le \tfrac{10N-1}{6N+3},\ N \ge 1 \tag{1.24}$$
$$H_{N,2} < 2 - \tfrac{1}{N},\ N \ge 2. \tag{1.25}$$

Hence, for $s > 2$

$$f(k,N,s) < N^{\frac{s}{2}-1}k^{-s}\left(\tfrac{\pi^2}{6} - \tfrac{N+1/2}{N^2+N+d}\right)^{-\frac{s}{2}},\ N \ge 1;$$
$$f(k,N,s) \le N^{\frac{s}{2}-1}k^{-s}\left(\tfrac{8}{5} - \tfrac{1}{N+\frac{2}{3}}\right)^{-\frac{s}{2}},\ N \ge 1;$$
$$f(k,N,s) \le N^{\frac{s}{2}-1}k^{-s}\left(\tfrac{13}{8} - \tfrac{1}{N+\frac{3}{5}}\right)^{-\frac{s}{2}},\ N \ge 1;$$
$$f(k,N,s) \le N^{\frac{s}{2}-1}k^{-s}\left(\tfrac{13}{8} - \tfrac{1}{N+\frac{2}{3}}\right)^{-\frac{s}{2}},\ N \ge 2,$$

and for $0 < s < 2$

$$f(k,N,s) > N^{\frac{s}{2}-1}k^{-s}\left(\tfrac{\pi^2}{6} - \tfrac{N+1/2}{N^2+N+1/3}\right)^{-\frac{s}{2}},\ N \ge 1;$$
$$f(k,N,s) \ge N^{\frac{s}{2}-1}k^{-s}\left(\tfrac{10N-1}{6N+3}\right)^{-\frac{s}{2}},\ N \ge 1;$$
$$f(k,N,s) \le N^{\frac{s}{2}-1}k^{-s}\left(2 - \tfrac{1}{N}\right)^{-\frac{s}{2}},\ N \ge 2.$$

1.1.3 Zipf law and Goldman inequality

From Goldman inequality we derived (1.6). For $q = 0,\ 0 < t < s$, (now $m = k/N,\ M = k$)

$$\frac{k^t - (\frac{k}{N})^t}{f(k,N,s)} - \frac{k^s - (\frac{k}{N})^s}{f(k,N,t)} \le N\left(k^t\left(\frac{k}{N}\right)^s - k^s\left(\frac{k}{N}\right)^t\right) \tag{1.26}$$

1° for $s > t = 1$ we have then

$$\frac{k - \frac{k}{N}}{f(k,N,s)} - \frac{k^s - (\frac{k}{N})^s}{f(k,N,1)} \le N\left(k\left(\frac{k}{N}\right)^s - k^s\frac{k}{N}\right)$$

i.e.

$$f(k,N,s) \ge \frac{1}{k^s}\cdot\frac{N^{s-1}(N-1)}{N - N^s + (N^s-1)H_N}. \tag{1.27}$$

Using (1.13)-(1.16) we get the following sequence of lower bounds for $f(k,N,s),\ s > 1$,

$$f(k,N,s) > \tfrac{1}{k^s}\cdot\tfrac{N^{s-1}(N-1)}{N-N^s+(N^s-1)\left(\gamma+\ln N+\frac{1}{2N}\right)},\ N > 1;$$
$$f(k,N,s) > \tfrac{1}{k^s}\cdot\tfrac{N^{s-1}(N-1)}{N-N^s+(N^s-1)\left(\gamma+\ln N+\frac{1}{2(N-1)}\right)},\ N > 1;$$

$$f(k,N,s) > \frac{1}{k^s}\cdot\frac{N^{s-1}(N-1)}{N-N^s+(N^s-1)\left(\gamma+\ln(N+1/2)+\frac{1}{24N^2}\right)},\ N>1;$$

$$f(k,N,s) > \frac{1}{k^s}\cdot\frac{N^{s-1}(N-1)}{N-N^s+(N^s-1)\left(\gamma+\ln(N+1/2)+\frac{1}{24(N+1/2)^2}-\frac{7}{960(N+1)^4}\right)}.\ N>1;$$

2° for $0<t<s=1$ in (1.26)

$$f(k,N,t) \le \frac{1}{k^t}\cdot\frac{N^{t-1}(N-1)}{N-N^t+(N^t-1)H_N}. \tag{1.28}$$

Using (1.13)-(1.16) we get the following sequence of upper bounds for $f(k,N,t),\ t<1$,

$$f(k,N,t) < \frac{1}{k^t}\cdot\frac{N^{t-1}(N-1)}{N-N^t+(N^t-1)\left(\gamma+\ln N+\frac{1}{2N}-\frac{1}{8N^2}\right)},\ N>1;$$

$$f(k,N,t) < \frac{1}{k^t}\cdot\frac{N^{s-1}(N-1)}{N-N^s+(N^s-1)\left(\gamma+\ln N+\frac{1}{2(N+1)}\right)},\ N>1;$$

$$f(k,N,t) < \frac{1}{k^t}\cdot\frac{N^{t-1}(N-1)}{N-N^t+(N^t-1)\left(\gamma+\ln(N+1/2)+\frac{1}{24(N+1)^2}\right)},\ N>1;$$

$$f(k,N,t) < \frac{1}{k^t}\cdot\frac{N^{t-1}(N-1)}{N-N^t+(N^t-1)\left(\gamma+\ln(N+1/2)+\frac{1}{24(N+1/2)^2}-\frac{7}{960(N+1)^4}\right)}\ N>1.$$

3° For $s>t=2$ in (1.26)

$$\frac{k^2-(\frac{k}{N})^2}{f(k,N,s)}-\frac{k^s-(\frac{k}{N})^s}{f(k,N,2)} \le N\left(k^2\left(\frac{k}{N}\right)^s-k^s\left(\frac{k}{N}\right)^2\right)$$

i.e.

$$f(k,N,s) \ge \frac{1}{k^s}\cdot\frac{N^{s-2}(N^2-1)}{N-N^{s-1}+(N^s-1)H_{N,2}},\ N>1. \tag{1.29}$$

Combining (1.30) with (1.20), (1.24) and (1.25) we get the sequence of inequalities

$$f(k,N,s) > \frac{1}{k^s}\cdot\frac{N^{s-2}(N^2-1)}{N-N^{s-1}+(N^s-1)\left(\frac{\pi^2}{6}-\frac{N+1/2}{N^2+N+1/3}\right)},\ N>1;$$

$$f(k,N,s) \ge \frac{1}{k^s}\cdot\frac{N^{s-2}(N^2-1)}{N-N^{s-1}+(N^s-1)\left(\frac{10N-1}{6N+3}\right)},\ N>1;$$

$$f(k,N,s) \ge \frac{1}{k^s}\cdot\frac{N^{s-2}(N^2-1)}{N-N^{s-1}+(N^s-1)\left(2-\frac{1}{N}\right)},\ N>2.$$

4° For $t>s=2$ in (1.26)

$$\frac{k^t-(\frac{k}{N})^t}{f(k,N,2)}-\frac{k^2-(\frac{k}{N})^2}{f(k,N,t)} \le N\left(k^t\left(\frac{k}{N}\right)^2-k^2\left(\frac{k}{N}\right)^t\right)$$

i.e.

$$f(k,N,t) \le \frac{1}{k^t} \cdot \frac{N^{t-2}(N^2-1)}{N - N^{t-1} + (N^t - 1)H_{N,2}}, \; N > 1. \tag{1.30}$$

Combining (1.30) with (1.20), (1.21), (1.22) and (1.23) we get the sequence of inequalities

$$\begin{aligned}
f(k,N,t) &< \tfrac{1}{k^t} \cdot \frac{N^{t-2}(N^2-1)}{N-N^{t-1}+(N^t-1)\left(\frac{\pi^2}{6}-\frac{N+1/2}{N^2+N+d}\right)}, \; N>1;\\
f(k,N,s) &\le \tfrac{1}{k^t} \cdot \frac{N^{t-2}(N^2-1)}{N-N^{t-1}+(N^t-1)\left(\frac{8}{5}-\frac{1}{N+\frac{2}{3}}\right)}, \; N>1;\\
f(k,N,t) &\le \tfrac{1}{k^t} \cdot \frac{N^{t-2}(N^t-1)}{N-N^{t-1}+(N^t-1)\left(\frac{13}{8}-\frac{1}{N+\frac{3}{5}}\right)}, \; N>1.\\
f(k,N,t) &\le \tfrac{1}{k^t} \cdot \frac{N^{t-2}(N^t-1)}{N-N^{t-1}+(N^t-1)\left(\frac{13}{8}-\frac{1}{N+\frac{2}{3}}\right)}, \; N\ge 2.
\end{aligned}$$

1.1.4 Further bounds via Lyapunov and Hölder inequality

Theorem 1.2 *For probability mass function* (1.39) *we have the following inequality, for* $0 < r < s < t$

$$\frac{[Nf(k,N,q,t)]^{-\frac{1}{t}} - [Nf(k,N,q,r)]^{-\frac{1}{r}}}{[Nf(k,N,q,t)]^{-\frac{1}{t}} - [Nf(k,N,q,s)]^{-\frac{1}{s}}} \le \frac{s(t-r)}{r(t-s)}. \tag{1.31}$$

Proof. Using Lyapunov inequality (see [14], p. 34, [15] p. 117). For $0 < r < s < t$

$$\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^s\right)^{t-r} \le \left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^r\right)^{t-s}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^t\right)^{s-r} \tag{1.32}$$

We can rewrite this as

$$[Nf(k,N,q,s)]^{-\frac{1}{s}} \le \left\{[Nf(k,N,q,r)]^{-\frac{1}{r}}\right\}^{\frac{r}{s}\frac{t-s}{t-r}}\left\{[Nf(k,N,q,t)]^{-\frac{1}{t}}\right\}^{\frac{t}{s}\frac{s-r}{t-r}} \tag{1.33}$$

Applying A-G inequality on right-hand side of (1.59) we have

$$[Nf(k,N,q,s)]^{-\frac{1}{s}} \le \frac{r}{s}\frac{t-s}{t-r}[Nf(k,N,q,r)]^{-\frac{1}{r}} + \frac{t}{s}\frac{s-r}{t-r}[Nf(k,N,q,t)]^{-\frac{1}{t}}$$

which we can rewrite as

$$\frac{[Nf(k,N,q,t)]^{-\frac{1}{t}} - [Nf(k,N,q,r)]^{-\frac{1}{r}}}{[Nf(k,N,q,t)]^{-\frac{1}{t}} - [Nf(k,N,q,s)]^{-\frac{1}{s}}} \le \frac{s(t-r)}{r(t-s)}.$$

□

Theorem 1.3 *For $\alpha > 1$, let (α,β) be a pair of Hölder conjugates. Then for $r,\ s > 0$ we have*

$$f(k,N,q,s+r) \geq f(k,N,q,s\alpha)^{\frac{1}{\alpha}} f(k,N,q,r\beta)^{\frac{1}{\beta}}. \tag{1.34}$$

Proof. Using Hölder inequality for sequences $\left\{\left(\frac{k+q}{i+q}\right)^r : i = 1,\ldots,N\right\}$ and $\left\{\left(\frac{k+q}{i+q}\right)^s : i = 1,\ldots,N\right\}$, we have

$$\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{r+s} \leq \left(\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{r\alpha}\right)^{1/\alpha}\left(\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{s\beta}\right)^{1/\beta}$$

i.e.

$$(f(k,N,q,s+r))^{-1} \leq f(k,N,q,s\alpha)^{-\frac{1}{\alpha}} f(k,N,q,r\beta)^{-\frac{1}{\beta}}.$$

□

Let

$$m = \begin{cases} \left(\frac{k+q}{N+q}\right)^{s-\frac{r\beta}{\alpha}}, & s\alpha > r\beta \\ \left(\frac{k+q}{1+q}\right)^{s-\frac{r\beta}{\alpha}}, & s\alpha < r\beta \end{cases} \tag{1.35}$$

and

$$M = \begin{cases} \left(\frac{k+q}{1+q}\right)^{s-\frac{r\beta}{\alpha}}, & s\alpha > r\beta \\ \left(\frac{k+q}{N+q}\right)^{s-\frac{r\beta}{\alpha}}, & s\alpha < r\beta. \end{cases} \tag{1.36}$$

Theorem 1.4 *For $\alpha > 1$, let (α,β) be a pair of Hölder conjugates. Then for $r,\ s > 0$ we have*

$$\frac{M-m}{f(k,N,q,s\alpha)} + \frac{mM^{\alpha} - Mm^{\alpha}}{f(k,N,q,r\beta)} \leq \frac{M^{\alpha} - m^{\alpha}}{f(k,N,q,r+s)}, \tag{1.37}$$

where m and M are defined with (1.35) *and* (1.36) *respectively.*

Proof. Follows from a conversion of the Hölder inequality and a discreet version of the linear functional in Theorem 4.14, [15], p. 114, applied for sequences

$$\left\{\left(\frac{k+q}{i+q}\right)^r : i = 1,\ldots,N\right\} \text{ and } \left\{\left(\frac{k+q}{i+q}\right)^s : i = 1,\ldots,N\right\}.$$

□

Another type of conversion of the Hölder inequality is given in [15], Theorem 4.16, p. 115. Similarly, as in the proof of Theorem 1.4, using discreet version of a linear functional, we get the next theorem.

Theorem 1.5 *Under the same assumptions as in Theorem 1.4, the following result holds*

$$f(k,N,q,r+s) \leq \frac{\alpha^{-\frac{1}{\alpha}}\beta^{-\frac{1}{\beta}}(M^{\alpha} - m^{\alpha})}{(M-m)^{\frac{1}{\alpha}}(mM^{\alpha} - Mm^{\alpha})^{\frac{1}{\beta}}}(f(k,N,q,s\alpha))^{\frac{1}{\alpha}}(f(k,N,q,r\beta))^{\frac{1}{\beta}}. \tag{1.38}$$

1.2 Analytical properties of Zipf-Mandelbrot law and Hurwitz $\zeta-$function

For $N \in \mathbb{N},\ q \geq 0,\ s > 0,\ k \in \{1,2,\ldots,N\}$, we can rewrite Zipf-Mandelbrot law (probability mass function) in the following form

$$f(k,N,q,s) = \frac{1/(k+q)^s}{\zeta(N,s,q)}, \tag{1.39}$$

where

$$\zeta(N,s,q) = \sum_{i=1}^{N} \frac{1}{(i+q)^s}, \tag{1.40}$$

$N \in \mathbb{N},\ q \geq 0,\ s > 0,\ k \in \{1,2,\ldots,N\}$. If total number of words N tends to infinity we denote

$$f(k,q,s) = \frac{1/(k+q)^s}{\zeta(s,q)}, \tag{1.41}$$

where

$$\zeta(s,q) = \sum_{i=1}^{\infty} \frac{1}{(i+q)^s} \tag{1.42}$$

we recognize as Hurwitz $\zeta-$function. This infinite case, when total mass is spread over all set of positive integers, particularly, is studied in [11]. Note here, that we use more suitable version of Hurwitz ζ function (see also [1]), since in the classical definition sum starts from zero and $q > 0$. However, this fact does not alter our conclusions about Hurwitz $\zeta-$function.

The are also quite different interpretation of Zipf-Mandelbrot law. As it is pointed out in [13] (see also [4], [18]), parameters in (1.39) can be interpreted in the following way: N is the number of species present and the parameters q and s have an ecological interpretation: q represents the diversity of the environment and s the predictability of the ecosystem, i.e. the average probability of the appearance of a species.

1.2.1 Monotonicity properties

As starting point, we use the next proposition on inequalities for sums of positive order ([14, pp. 36] , [15, pp. 165]).

Proposition 1.2 *If* $a_i \geq 0,\ i \in \mathbb{N}$ *then for* $0 < t < s$

$$\left(\sum_{i=1}^{\infty} a_i^s\right)^{\frac{1}{s}} \leq \left(\sum_{i=1}^{\infty} a_i^t\right)^{\frac{1}{t}}. \tag{1.43}$$

Theorem 1.6

i) The function $s \mapsto [\zeta(N,s,q)]^{1/s}$ *is decreasing i.e. for* $s > t > 0$

$$[\zeta(N,s,q)]^{1/s} \leq [\zeta(N,t,q)]^{1/t}.$$

ii) The function $s \mapsto [f(k,N,q,s)]^{1/s}$ *is increasing i.e. for* $s > t > 0$

$$[f(k,N,q,s)]^{1/s} \geq [(f(k,N,q,t)]^{1/t}.$$

iii) The function $s \mapsto [\zeta(s,q)]^{1/s}$ *is decreasing i.e. for* $s > t > 0$

$$[\zeta(s,q)]^{1/s} \leq [\zeta(t,q)]^{1/t}.$$

iv) The function $s \mapsto [f(k,q,s)]^{1/s}$ *is increasing i.e. for* $s > t > 0$

$$[f(k,q,s)]^{1/s} \geq [(f(k,q,t)]^{1/t}.$$

Proof.
i) We use the Proposition 1.2, for

$$a_i = \begin{cases} \frac{1}{i+q}, & i = 1,\ldots,N; \\ 0, & i > N. \end{cases}$$

ii) Follows from i)-part and

$$\frac{1}{f(k,N,q,s)} = \sum_{i=1}^{N} \left(\frac{k+q}{i+q}\right)^s = (k+q)^s \zeta(N,s,q). \tag{1.44}$$

iii) Use Proposition 1.2 for $a_i = \frac{1}{i+q}$, $i \in \mathbb{N}$.
iv) Follows from iii)-part and

$$\frac{1}{f(k,q,s)} = (k+q)^s \zeta(s,q). \tag{1.45}$$

□

Theorem 1.7 *The function*

$$s \mapsto (Nf(k,N,q,s))^{1/s} \tag{1.46}$$

is decreasing i.e. for $s > t > 0$

$$(Nf(k,N,q,s))^{1/s} \leq (Nf(k,N,q,t))^{1/t}. \tag{1.47}$$

Proof. From (1.44) it follows

$$\frac{1}{Nf(k,N,q,s)} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{s}, \tag{1.48}$$

i.e.

$$(Nf(k,N,q,s))^{-1/s} = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{s}\right]^{1/s}. \tag{1.49}$$

Denote $x_i = \frac{k+q}{i+q},\ i = 1,\ldots,N$. Then the right-hand side of (1.49) is the power mean

$$M_N^{[s]}(x_{\overline{1,N}}) := \left[\frac{1}{N}\sum_{i=1}^{N} x_i^{\,s}\right]^{1/s}.$$

Using well-known fact, that $s \mapsto M_N^{[s]}(x_{\overline{1,N}})$ is increasing function (see for example [14, 15]) we conclude that the function

$$s \mapsto (Nf(k,N,q,s))^{1/s} \tag{1.50}$$

is decreasing. □

1.2.2 Log-convexity and exponential convexity

Let us recall well-known Lyapunov inequality, for sequences ([14, pp. 34], [15, pp. 117]).

Proposition 1.3 *If* $a_i \geq 0,\ i \in \mathbb{N}$, *then for* $0 < r < s < t$

$$\left(\sum_{i=1}^{\infty} a_i^s\right)^{t-r} \leq \left(\sum_{i=1}^{\infty} a_i^r\right)^{t-s}\left(\sum_{i=1}^{\infty} a_i^t\right)^{s-r}. \tag{1.51}$$

If we set $a_i = \frac{1}{i+q},\ i \in \mathbb{N}$ in (1.51) we get

Corollary 1.1 *For* $1 < r < s < t$

$$\zeta^{t-r}(s,q) \leq \zeta^{t-s}(r,q)\zeta^{s-r}(t,q). \tag{1.52}$$

In the next theorem we prove, log-concavity of $s \mapsto f(k,N,q,s)$ and log-convexity of $s \mapsto \zeta(s,q)$.

Theorem 1.8 *Let* $\lambda \in (0,1)$.

i) For $0 < r < t$,

$$\zeta(N,\lambda r + (1-\lambda)t,q) \leq \zeta^{\lambda}(N,r,q)\zeta^{1-\lambda}(N,t,q).$$

ii) For $0 < r < t$,

$$\left(f(k,N,q,\lambda r + (1-\lambda)t)\right)^{-1} \leq \left(f(k,N,q,r)\right)^{-\lambda} \left(f(k,N,q,t)\right)^{-(1-\lambda)}.$$

iii) For $1 < r < t$,

$$\zeta(\lambda r + (1-\lambda)t, q) \leq \zeta^{\lambda}(r,q)\zeta^{1-\lambda}(t,q).$$

iv) For $1 < r < t$,

$$\left(f(k,q,\lambda r + (1-\lambda)t)\right)^{-1} \leq \left(f(k,q,r)\right)^{-\lambda} \left(f(k,q,t)\right)^{-(1-\lambda)}.$$

Proof.
i) For $0 < r < t$ and $\lambda \in (0,1)$ we set

$$a_i = \begin{cases} \frac{1}{i+q}, & i = 1,\ldots,N; \\ 0, & i > N. \end{cases}$$

and $s = \lambda r + (1-\lambda)t$ in (1.51):

$$\left(\sum_{i=1}^{N}\left(\frac{1}{i+q}\right)^{\lambda r+(1-\lambda)t}\right)^{t-r} \leq \left(\sum_{i=1}^{N}\left(\frac{1}{i+q}\right)^{r}\right)^{\lambda(t-r)} \left(\sum_{i=1}^{N}\left(\frac{1}{i+q}\right)^{t}\right)^{(1-\lambda)(t-r)}.$$

ii) Follows from (1.44) and i)-part.
iii) We set $a_i = \frac{1}{i+q}$ and $s = \lambda r + (1-\lambda)t$ in (1.51).
iv) Follows from iii)-part and (1.45). □

We can conclude even more since this result can be extended to exponential convexity [5].

Definition 1.1 *A function* $h : I \to \mathbb{R}$ *is exponentially convex on an interval* $I \subseteq \mathbb{R}$ *if it is continuous and*

$$\sum_{i,j=1}^{n} \xi_i \xi_j h\left(\frac{x_i + x_j}{2}\right) \geq 0$$

for all $n \in \mathbb{N}$ *and all choices* $\xi_i \in \mathbb{R},\ x_i \in I,\ i = 1,\ldots,n$.

Theorem 1.9 *The function* $s \mapsto \zeta(s,q)$ *is exponentially convex function on* $(1,\infty)$.

Proof. For a given $n \in \mathbb{N}$ let $\xi_m \in \mathbb{R},\ s_m \in (1,\infty)$ $(m = 1,\ldots,n)$ we have

$$\sum_{l,m=1}^{n} \xi_l \xi_m \zeta\left(\frac{s_l + s_m}{2}, q\right) = \sum_{l,m=1}^{n} \xi_l \xi_m \sum_{i=1}^{\infty} \frac{1}{(i+q)^{\frac{s_l+s_m}{2}}} \tag{1.53}$$

$$= \sum_{i=1}^{\infty} \sum_{l,m=1}^{n} \xi_l \xi_m \frac{1}{(i+q)^{\frac{s_l+s_m}{2}}} \tag{1.54}$$

$$= \sum_{i=1}^{\infty} \left(\sum_{m=1}^{n} \frac{1}{(i+q)^{\frac{s_m}{2}}}\right)^2 \geq 0. \tag{1.55}$$

Since the function $s \mapsto \zeta(s,q)$ is continuous function on $(1,\infty)$, we conclude its exponential convexity on $(1,\infty)$. □

Using (1.45) we have also the next corollary.

Corollary 1.2 *The function* $s \mapsto (f(k,q,s))^{-1}$ *is exponentially convex function on* $(1,\infty)$.

Proof. This is consequence of (1.45) and the fact that exponential convexity is closed under finite multiplication of exponentially convex functions. □

Corollary 1.3 *The matrices* $\left[\left(\zeta(\frac{s_l+s_m}{2},q)\right)\right]_{l,m=1}^{n}$ *and* $\left[\left(f(k,q,\frac{s_l+s_m}{2})\right)^{-1}\right]_{l,m=1}^{n}$ *are positive semi definite for all* $n \in \mathbb{N},\ s_1,\dots,s_n$ *in* $(1,\infty)$.

We can also deduce exponential convexity from diversity point of view, notion mentioned in the introduction.

Theorem 1.10 *For any* $s>0,\ N\in\mathbb{N}$, *the function*

$$q \mapsto \zeta(N,s,q)$$

is exponentially convex on $(0,\infty)$.

Proof. For $k=1,\dots,N$, using the Laplace transform,

$$\frac{1}{(k+q)^s} = \int_0^\infty e^{-(k+q)t}\frac{t^{s-1}}{\Gamma(s)}dt$$

and the fact

$$\sum_{i,j=1}^{n} \xi_i\xi_j \exp\left[-\left(k+\frac{q_i+q_j}{2}\right)t\right] = e^{-kt}\left(\sum_{i=1}^{n}\xi_i \exp\left(-\frac{q_i}{2}t\right)\right)^2 \ge 0,$$

we conclude exponential convexity of the function $q \mapsto \frac{1}{(k+q)^s}$ on $(0,\infty)$. Now $q \mapsto \zeta(N,s,q)$ is exponentially convex on $(0,\infty)$ as a finite sum of exponentially convex functions. □

Theorem 1.11 *For any* $s>1$, *the function*

$$q \mapsto \zeta(s,q)$$

is exponentially convex on $(0,\infty)$.

Proof. Using Mellin transformation

$$\zeta(s,q) = \frac{1}{\Gamma(s)}\int_0^\infty \frac{t^{s-1}e^{-(q+1)t}}{1-e^{-t}}dt$$

and

$$\sum_{i,j=1}^{n} \xi_i\xi_j \exp\left(-\left(\frac{q_i+q_j}{2}+1\right)t\right) = \left(\sum_{i=1}^{n}\xi_i \exp\left(-\frac{q_i+1}{2}t\right)\right)^2 \ge 0,$$

we conclude exponential convexity of $q \mapsto \zeta(s,q)$ on $(0,\infty)$. □

Corollary 1.4 *For $s > 1$, the matrix $\left[\zeta\left(s, \left(\frac{q_l+q_m}{2}\right)\right)\right]_{l,m=1}^{n}$ is positive semi definite for all $n \in \mathbb{N},\ q_1, \ldots, q_n \in (0,\infty)$.*

Corollary 1.5 *For any $s > 1$, the function*

$$q \mapsto \zeta(s,q)$$

is log-convex on $(0,\infty)$.

1.2.3 Log subadditivity

Let us recall Chebyshev's inequality (see [14, pp. 27], [15, pp. 197]).

Theorem 1.12 *Let $(a_1,\ldots,a_N)$ and $(b_1,\ldots,b_N)$ be two $N-$tuples of real numbers such that*

$$(a_i - a_j)(b_i - b_j) \geq 0,\ for\ i,j = 1,\ldots,N,$$

and $(w_1,\ldots,w_N)$ be a positive $n-$tuple. Then

$$\left(\sum_{i=1}^{N} w_i\right)\left(\sum_{i=1}^{N} w_i a_i bi\right) \geq \left(\sum_{i=1}^{n} w_i a_i\right)\left(\sum_{i=1}^{N} w_i b_i\right). \tag{1.56}$$

Theorem 1.13 *The function $s \mapsto Nf(k,N,q,s)$ is log subadditive, i.e. for $s,r > 0$*

$$Nf(k,N,q,s+r) \leq [Nf(k,N,q,s)]\,[Nf(k,N,q,r)]\,. \tag{1.57}$$

Proof. We apply Chebyshev's inequality (1.79) for

$$a_i = \left(\frac{k+q}{i+q}\right)^s,\ b_i = \left(\frac{k+q}{i+q}\right)^r,\ w_i = \frac{1}{N};\ i = 1,\ldots,N.$$

Hence we get

$$\frac{1}{N}\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{s+r} \geq \left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{s}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{r}\right)$$
$$\Rightarrow \frac{1}{Nf(k,N,q,s+r)} \geq \frac{1}{Nf(k,N,q,s)}\frac{1}{Nf(k,N,q,r)},$$

concluding (1.81). □

Theorem 1.14 *The function $u \mapsto \left[f(k,N,q,u^{-1})\right]^{-u}$ is log-convex.*

Proof. Using Lyapunov inequality in Proposition 1.3, for $0 < r < s < t$

$$\left(\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{s}\right)^{t-r} \leq \left(\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{r}\right)^{t-s}\left(\sum_{i=1}^{N}\left(\frac{k+q}{i+q}\right)^{t}\right)^{s-r}. \tag{1.58}$$

Using (1.49) we rewrite this as

$$[f(k,N,q,s)]^{-\frac{1}{s}} \leq \left\{[f(k,N,q,r)]^{-\frac{1}{r}}\right\}^{\frac{r}{s}\frac{t-s}{t-r}} \left\{[f(k,N,q,t)]^{-\frac{1}{t}}\right\}^{\frac{t}{s}\frac{s-r}{t-r}} \tag{1.59}$$

Now we substitute $t = 1/x,\ r = 1/y,\ \lambda = \frac{t}{s}\frac{s-r}{t-r}$ in (1.59), and since $1-\lambda = \frac{r}{s}\frac{t-s}{t-r}$, $s = [\lambda x + (1-\lambda)y]^{-1}$, we have

$$\left[f(k,N,q,[\lambda x+(1-\lambda)y]^{-1})\right]^{-[\lambda x+(1-\lambda)y]} \leq \left\{\left[f(k,N,q,x^{-1})\right]^{-x}\right\}^{\lambda} \left\{\left[f(k,N,q,y^{-1})\right]^{-y}\right\}^{1-\lambda},$$

concluding log-convexity of the function $u \mapsto \left[f(k,N,q,u^{-1})\right]^{-u}$. □

1.2.4 Gini means and further monotonicity

For positive $n-$tuple $(a_1,\ldots,a_n),\ \alpha,\beta \in \mathbb{R}$, Gini means are defined with

$$G(\alpha,\beta) = \begin{cases} \left(\dfrac{\sum_{i=1}^{n} a_i^{\alpha}}{\sum_{i=1}^{n} a_i^{\beta}}\right)^{\frac{1}{\alpha-\beta}}, & \alpha \neq \beta; \\ \exp\left(\sum_{i=1}^{n} a_i^{\alpha}\ln a_i / \sum_{i=1}^{n} a_i^{\alpha}\right), & \alpha = \beta. \end{cases} \tag{1.60}$$

It is known then see [15, pp. 119],

$$G(\alpha_1,\beta_1) \leq G(\alpha_2,\beta_2), \tag{1.61}$$

for $\alpha_1 \leq \alpha_2,\ \beta_1 \leq \beta_2,\ \alpha_1 \neq \beta,\ \alpha_2 \neq \beta_2$.
If we choose $a_i = \frac{k+q}{i+q}$ in (1.60) we will get Zip-Mandelbrot means:

$$Z(\alpha,\beta) = \begin{cases} \left(\frac{f(k,N,q,\beta)}{f(k,N,q,\alpha)}\right)^{\frac{1}{\alpha-\beta}}, & \alpha \neq \beta; \\ \left[(k+q)^{\alpha}\zeta(N,s,\alpha)\right]^{\frac{(k+q)^{\alpha}H_{N,q,\alpha}}{\alpha f(k,N,q,\alpha)}} \exp\left(-\frac{(k+q)^{\alpha}}{\alpha f(k,N,q,\alpha)}E(k,N,q,\alpha)\right), & \alpha = \beta. \end{cases} \tag{1.62}$$

where

$$E(k,N,q,\alpha) = -\sum_{k=1}^{N} f(k,N,q,\alpha)\ln f(k,N,q,\alpha)$$

denotes Shannon entropy of the law (1.39) (for related results see also [9]).
Using (1.86) we can now formulate the next theorem.

Theorem 1.15 *For* $0 < \alpha_1 \leq \alpha_2,\ 0 < \beta_1 \leq \beta_2,\ \alpha_1 \neq \beta,\ \alpha_2 \neq \beta_2$;

$$Z(\alpha_1,\beta_1) \leq Z(\alpha_2,\beta_2). \tag{1.63}$$

The expectation of the Zipf-Mandelbrot law is

$$\sum_{k=1}^{N} kf(k,N,q,s) = \frac{1}{\zeta(N,s,q)} \sum_{k=1}^{N} \frac{k+q-q}{(k+q)^s} = \frac{\zeta(N,s-1,q)}{\zeta(N,s,q)} - q.$$

This is a decreasing function over s, as the next theorem shows.

Theorem 1.16 *The function*

$$s \mapsto \frac{\zeta(N,s-1,q)}{\zeta(N,s,q)}$$

is decreasing on $\mathbb{R}_+$.

Proof. We set $a_i = \frac{1}{i+q}$, $i = 1,\ldots,N$ and $\alpha = s-1$, $\beta = s$ in (1.60). According (1.86), for $0 < s < t$, we have

$$\left(\frac{\zeta(N,s-1,q)}{\zeta(N,s,q)}\right)^{-1} \leq \left(\frac{\zeta(N,t-1,q)}{\zeta(N,t,q)}\right)^{-1}.$$

□

Of course, result can be extended to Hurwitz $\zeta-$function.

Corollary 1.6 *The function*

$$s \mapsto \frac{\zeta(s-1,q)}{\zeta(s,q)}$$

is decreasing on $\mathbb{R}_+$.

Remark 1.2 General remark in this section is that parameters α, β in (1.85) could be any real numbers, so Theorems 1.27 and 1.16 are also valid on $\mathbb{R}^2$ and $\mathbb{R}$, respectively.

1.3 Hybrid Zipf-Mandelbrot law

There is a unified approach, maximization of Shannon entropy, that naturally follows the path of generalization from Zipf's to hybrid Zipf's law. Extending this idea, in this section, we make transition from Zipf-Mandelbrot to hybrid Zipf-Mandelbrot law. It is interesting that examination of its densities provides some new insights of Lerch's transcendent (see [8]).

1.3.1 Shannon entropy and Zipf-Mandelbrot law

Here we extend use the maximum entropy approach in [17] to Zipf's law in order to deduce Zipf-Mandelbrot law, i.e. we maximize

$$S = -\sum_{i\in I} p_i \ln p_i \tag{1.64}$$

subject to some constraints. Trivial constraint is of course $\sum\limits_{i\in I} p_i = 1$.

Theorem 1.17 *Let $I = \{1,\dots,N\}$ or $I = \mathbb{N}$. For a given $q \geq 0$ and $\chi \geq 0$, a probability distribution, concentrated on I, that maximizes Shannon entropy under additional constraint*

$$\sum_{k\in I} p_k \ln(k+q) = \chi \tag{1.65}$$

is Zipf-Mandelbrot law.

Proof. If $I = \{1,\dots,N\}$, in a very standard procedure, we set two Lagrange multipliers λ and s and consider expression

$$\hat{S} = -\sum_{k=1}^{N} p_k \ln p_k - \lambda\left(\sum_{k=1}^{N} p_k - 1\right) - s\left(\sum_{k=1}^{N} p_k \ln(k+q) - \chi\right).$$

Just for convenience we can, of course, replace $\lambda \longleftrightarrow \ln\lambda - 1$, and now conider

$$\hat{S} = -\sum_{k=1}^{N} p_k \ln p_k - (\ln\lambda - 1)\left(\sum_{k=1}^{N} p_k - 1\right) - s\left(\sum_{k=1}^{N} p_k \ln(k+q) - \chi\right)$$

instead.
From $\hat{S}_{p_k} = 0,\ k = 1,\dots,N$ we deduce

$$p_k = \frac{1}{\lambda(k+q)^s},$$

and combining this with $\sum\limits_{k=1}^{N} p_k = 1$, we have

$$\lambda = \sum_{k=1}^{N} \frac{1}{(k+q)^s},$$

where $s > 0$, concluding

$$p_k = \frac{1/(k+q)^s}{\zeta(N,s,q)},\ k = 1,\dots,N.$$

The case $I = \mathbb{N}$ is treated in a similar manner with the restriction $s > 1$:

$$p_k = \frac{1/(k+q)^s}{\zeta(s,q)},\ k \in \mathbb{N}.$$

□

Remark 1.3

(i) If X is the random variable with values at I and probability law $(p_i,\ i \in I)$, then χ from (1.65) is in fact expectation of the random variable $\ln(X+q)$, which depends on X.

(ii) Observe here that for Zipf-Mandelbrot law (1.39) Shannon entropy (1.64) can be bounded from above (see [12]):

$$S = -\sum_{k=1}^{\infty} f(k,q,s)\ln f(k,q,s) \leq -\sum_{k=1}^{\infty} f(k,q,s)\ln q_k, \tag{1.66}$$

where $(q_k :\ k \in \mathbb{N})$ is any sequence of positive numbers such that $\sum\limits_{k=1}^{\infty} q_k = 1$.

1.3.2 Hybrid Zipf-Mandelbrot law

The same technique of maximum entropy we apply with one additional constraint. The derived probability law we will call *hybrid Zipf-Mandelbrot law.*

Theorem 1.18 *Let $I = \{1,\dots,N\}$ or $I = \mathbb{N}$. For a given $q \geq 0$, $\chi \geq 0$ and $\mu \geq 0$, a probability distribution, concentrated on I, that maximizes Shannon entropy under additional constraints*

$$\sum_{k\in I} p_k \ln(k+q) = \chi,\ \sum_{k\in I} kp_k = \mu$$

is hybrid Zipf-Mandelbrot law:

$$p_k = \frac{w^k}{(k+q)^s \Phi^*(s,q,w)},\ k \in I,$$

where

$$\Phi_I^*(s,q,w) = \sum_{k\in I} \frac{w^k}{(k+q)^s}.$$

Proof. We consider first $I = \{1,\dots,N\}$ and then we maximize

$$\hat{S} = -\sum_{k=1}^{N} p_k \ln p_k + \ln w\left(\sum_{k=1}^{N} kp_k - \mu\right) - (\ln\lambda - 1)\left(\sum_{k=1}^{N} p_k - 1\right) - s\left(\sum_{k=1}^{N} p_k \ln(k+q) - \chi\right).$$

$\hat{S}_{p_k} = 0,\ k = 1,\dots,N$ gives us

$$-\ln p_k + k\ln w - \ln\lambda - s\ln(k+q) = 0,$$

i.e.

$$p_k = \frac{w^k}{\lambda(k+q)^s}.$$

Using $\sum\limits_{k=1}^{N} p_k = 1$, we get $\lambda = \sum\limits_{k=1}^{N} \frac{w^k}{(k+q)^s}$ and we recognize this as the partial sum of Lerch's transcendent

$$\Phi_N^*(s,q,w) = \sum_{k=1}^{N} \frac{w^k}{(k+q)^s},$$

with $w \geq 0, s > 0$.
In the infinite case $I = \mathbb{N}$ we have restrictions either $w < 1, s > 0$ or $w = 1, s > 1$ and

$$\lambda = \sum_{k=1}^{\infty} \frac{w^k}{(k+q)^s}$$

we recognize as Lerch's transcendent that we will denote with $\Phi^*(s,q,w)$. □

∴

Let us denote

$$f_h(w,N,k,q,s) = \frac{w^k}{(k+q)^s \Phi_N^*(s,q,w)}, \; k = 1,\ldots,N \tag{1.67}$$

and

$$f_h(w,k,q,s) = \frac{w^k}{(k+q)^s \Phi^*(s,q,w)}, \tag{1.68}$$

hybrid Zipf-Mandelbrot law on finite and infinite state space, respectively.

Remark 1.4 Some remarks are needed.

(i) Observe that constraint with the μ is in fact the expectation of the law.

(ii) There is a slight difference between Lerch's transcendent defined in [2] p. 27 and with our understanding of Lerch's transcendent: we don't have 0th summand.

(iii) We omitted the full bordered Hessian discussion in proofs of Theorems 1.17 and 1.18 as mere standard procedure.

(iv) Observe, further, that for hybrid Zipf-Mandelbrot law (1.68) Shannon entropy (1.64) can be bounded from above (see [12]):

$$S = -\sum_{k=1}^{\infty} f_h(k,q,s) \ln f_h(k,q,s) \leq -\sum_{k=1}^{\infty} f_h(k,q,s) \ln q_k, \tag{1.69}$$

where $(q_k : \; k \in \mathbb{N})$ is any sequence of positive numbers such that $\sum_{k=1}^{\infty} q_k = 1$.

1.3.3 Properties of the hybrid Zipf-Mandelbrot law

Now we examine analytical properties of the Lerch's transcendent and the hybrid Zipf-Mandelbrot law.

Theorem 1.19 *The functions*

$$s \mapsto \left(\frac{w-w^{N+1}}{w^k-w^{k+1}} f_h(w,N,k,q,s)\right)^{1/s} \tag{1.70}$$

and

$$s \mapsto \left(\frac{w}{w^k-w^{k+1}} f_h(w,k,q,s)\right)^{1/s} \tag{1.71}$$

are decreasing on $(0,\infty)$.

Proof. From (1.67) it follows

$$\frac{1}{f_h(w,N,k,q,s)} = \frac{1}{w^k}\sum_{i=1}^{N} w^i \left(\frac{k+q}{i+q}\right)^s$$

i.e.

$$\left(\frac{w^k-w^{k+1}}{(w-w^{N+1})h(w,N,k,q,s)}\right)^{1/s} = \left(\frac{1}{\frac{w-w^{N+1}}{1-w}}\sum_{i=1}^{N} w^i \left(\frac{k+q}{i+q}\right)^s\right)^{1/s}. \tag{1.72}$$

The right-hand side of (1.72) is power mean, which is increasing function on parameter s. □

Now we recall well-known Lyapunov inequality, for isotonic functionals (for details see [15, pp. 117]): for $0<r<s<t$

$$A(g^s)^{t-r} \le A(g^r)^{t-s} A(g^t)^{s-r}. \tag{1.73}$$

Theorem 1.20

i) For $N\in\mathbb{N},\ w>0,\ q\ge 0,\ 0<r<s<t,$

$$[\Phi_N^*(s,q,w)]^{t-r} \le [\Phi_N^*(s,q,w)]^{t-s}\,[\Phi_N^*(s,q,w)]^{s-r}.$$

ii) For $0<w<1,\ q\ge 0,\ 1<r<s<t,$

$$[\Phi^*(s,q,w)]^{t-r} \le [\Phi^*(s,q,w)]^{t-s}\,[\Phi^*(s,q,w)]^{s-r}.$$

Proof. i) We apply (1.73) to the linear functional

$$A(g) = \sum_{k=1}^{N} w^k g(k)$$

and replace $g(k)=\frac{1}{k+q}$.
ii) Similarly, if we define

$$A(g)=\sum_{k=1}^{\infty} w^k g(k)$$

then the result follows from (1.73), if we choose $g(k)=\frac{1}{k+q}$. □

We can now conclude log-convexity of Lerch's transcendent and log-concavity of hybrid Zipf-Mandelbrot law.

Corollary 1.7 *Let $\lambda \in (0,1)$.*

i) For $0<r<t$, $N\in\mathbb{N}$, $w>0$, $q\geq 0$, $0<r<s<t$,

$$\Phi_N^*(\lambda r+(1-\lambda)t,q,w)\leq [\Phi_N^*(r,q,w)]^{\lambda}\,[\Phi_N^*(t,q,w)]^{1-\lambda}.$$

ii) For $1<r<t$, $0<w<1$, $q\geq 0$

$$\Phi^*(\lambda r+(1-\lambda)t,q,w)\leq [\Phi^*(r,q,w)]^{\lambda}\,[\Phi^*(t,q,w)]^{1-\lambda}.$$

iii) For $N\in\mathbb{N}$, $w>0$, $q\geq 0$, $0<r<s<t$,

$$(f_h(w,N,k,q,\lambda r+(1-\lambda)t))^{-1}\leq (f_h(w,N,k,q,r))^{-\lambda}\,(f_h(w,N,k,q,t))^{-(1-\lambda)}.$$

iv) For $0<w<1$, $q\geq 0$, $1<r<s<t$,

$$(f_h(w,k,q,\lambda r+(1-\lambda)t))^{-1}\leq (f_h(w,k,q,r))^{-\lambda}\,(f_h(w,k,q,t))^{-(1-\lambda)}.$$

Proof. i) and ii) follow from Theorem 6.89. iii) and iv) follow from (1.67) and (1.68) respectively. □

The results in the previous corollary can be extended further to exponential convexity.

Definition 1.2 *A function $h: I\to\mathbb{R}$ is exponentially convex on an open interval $I\subseteq\mathbb{R}$ if it is continuous and*

$$\sum_{i,j=1}^{n} \xi_i\xi_j h\left(\frac{x_i+x_j}{2}\right)\geq 0$$

for all $n\in\mathbb{N}$ and all choices $\xi_i\in\mathbb{R}$, $x_i\in I$, $i=1,\ldots,n$.

Theorem 1.21 *The function $s\mapsto \Phi^*(s,q,w)$ is exponentially convex function on $(1,\infty)$.*

Proof. For a given $n\in\mathbb{N}$ let $\xi_m\in\mathbb{R}$, $s_m\in(1,\infty)$ $(m=1,\ldots,n)$ we have

$$\sum_{l,m=1}^{n} \xi_l\xi_m\Phi^*(\frac{s_l+s_m}{2},q,w) = \sum_{l,m=1}^{n} \xi_l\xi_m \sum_{i=1}^{\infty} \frac{w^i}{(i+q)^{\frac{s_l+s_m}{2}}} \tag{1.74}$$

$$= \sum_{i=1}^{\infty} w^i \sum_{l,m=1}^{n} \frac{\xi_l\xi_m}{(i+q)^{\frac{s_l+s_m}{2}}} \tag{1.75}$$

$$= \sum_{i=1}^{\infty} w^i \left(\sum_{m=1}^{n} \frac{\xi_m}{(i+q)^{\frac{s_m}{2}}} \right)^2 \geq 0. \tag{1.76}$$

Since the function $s \mapsto \Phi^*(s,q,w)$ is continuous function on $(1,\infty)$, we conclude its exponential convexity on $(1,\infty)$. □

Using (1.68) we have also the next corollary.

Corollary 1.8 *The function* $s \mapsto (f_h(w,k,q,s))^{-1}$ *is exponentially convex function on* $(1,\infty)$.

Corollary 1.9 *The matrices* $\left[\left(\Phi^*(\frac{s_l+s_m}{2},q,w)\right)\right]_{l,m=1}^{n}$ *and* $\left[\left(f_h(w,k,q\frac{s_l+s_m}{2})\right)^{-1}\right]_{l,m=1}^{n}$ *are positive semidefinite for all* $n \in \mathbb{N}$, $s_1,\ldots,s_n$ *in* $(1,\infty)$.

We can deduce exponential convexity for the second parameter in generalized polylogarithm function. First, we will prove theorem on integral representation of generalized polylogarithm function as a variant of Mellin transformation for Hurwitz ζ function.

Lemma 1.1 *For* $0 < w < 1$, $q \geq 0$

$$\Phi^*(s,q,w) = \frac{w}{\Gamma(s)} \int_0^{\infty} \frac{u^{s-1}e^{-(q+1)u}}{1-we^{-u}} du. \tag{1.77}$$

Proof. In Gamma function integral we change variable, $x = (n+q)u$,

$$\Gamma(s) = \int_0^{\infty} e^{-x}x^{s-1}dx = (k+q)^s \int_0^{\infty} e^{-(k+q)u}u^{s-1}du,$$

hence,

$$\Rightarrow (k+q)^{-s}\Gamma(s) = \int_0^{\infty} e^{-ku}e^{-qu}u^{s-1}du. \tag{1.78}$$

By multiplying both sides of (1.78) with w^k , summing over $k \in \mathbb{N}$, and using Beppo-Levi's theorem on the right side, we have

$$\begin{aligned}\Phi^*(s,q,w)\Gamma(s) &= \int_0^{\infty} \sum_{k=1}^{\infty} w^k e^{-ku}e^{-qu}u^{s-1}du \\ &= w\int_0^{\infty} \frac{u^{s-1}e^{-(q+1)u}}{1-we^{-u}}du.\end{aligned}$$

□

Theorem 1.22 *The function* $q \mapsto \Phi^*(s,q,w)$ *is exponentially convex function on* $(0,\infty)$.

Proof. For a given $n \in \mathbb{N}$, $\xi_m \in \mathbb{R}$, $q_m \in (0,\infty)$ $(m = 1,\ldots,n)$, we have, using (1.77) and

$$\sum_{i,j=1}^{n} \xi_i\xi_j \exp\left(-\left(\frac{q_i+q_j}{2}+1\right)t\right) = \left(\sum_{i=1}^{n} \xi_i \exp\left(-\frac{q_i+1}{2}t\right)\right)^2 \geq 0,$$

concluding

$$\sum_{i,j=1}^{n} \xi_i \xi_j \Phi^*(s, \frac{q_i+q_j}{2}, w) \geq 0.$$

□

Corollary 1.10 *For $s > 1$, the matrix $\left[\Phi^*(s, \frac{q_i+q_j}{2}, w)\right]_{i,j=1}^{n}$ is positive semi definite for all $n \in \mathbb{N}$, $q_1, \ldots, q_n$ in $(0,\infty)$.*

Corollary 1.11 *For any $s > 1$, the function*

$$q \mapsto \Phi^*(s,q,w)$$

is log-convex on $(0,\infty)$.

Theorem 1.23 *The function $w \mapsto \frac{\Phi^*(s,q,w)}{w}$ is exponentially convex on $(0,1)$.*

Proof. From $\frac{1}{w} = \int_0^\infty e^{-wt}dt$ we have

$$\frac{1}{1-we^{-u}} = \int_0^\infty e^{t+we^{-u}t}dt.$$

If we now rewrite (1.77)

$$\frac{\Phi^*(s,q,w)}{w} = \frac{1}{\Gamma(s)} \int_0^\infty u^{s-1} e^{-(q+1)u} \int_0^\infty e^{t+we^{-u}t} dt du,$$

and use Fubini with

$$\sum_{i,j=1}^{n} \xi_i \xi_j e^{t+\frac{w_i+w_j}{2}e^{-u}t} = e^t \left(\sum_{i=1}^{n} e^{\frac{w_i e^{-u}t}{2}} \right)^2 \geq 0,$$

our proof is done. □

Corollary 1.12 *For any $\alpha > 1$ the function $w \mapsto \frac{\Phi^*(s,q,w)}{w^\alpha}$ is exponentially convex on $(0,1)$.*

Proof. This follows from the fact that, for $\gamma > 0$, $x \mapsto x^{-\gamma}$ is exponentially convex on $(0,1)$ and that product of exponentially convex function (on the same domain) is again exponentially convex (for details see [5]). □

Let us recall Chebyshev's inequality (see [15, pp. 197]).

Theorem 1.24 *Let $(a_1,\ldots,a_N)$ and $(b_1,\ldots,b_N)$ be two N-tuples of real numbers such that*

$$(a_i - a_j)(b_i - b_j) \geq 0, \; for \; i,j = 1,\ldots,N,$$

and $(w_1,\ldots,w_N)$ be a positive $n-$tuple. Then

$$\left(\sum_{i=1}^{N} w_i\right)\left(\sum_{i=1}^{N} w_i a_i bi\right) \geq \left(\sum_{i=1}^{N} w_i a_i\right)\left(\sum_{i=1}^{N} w_i b_i\right). \tag{1.79}$$

Remark 1.5 The previous theorem can be extended to infinite sequences if we impose some obvious convergence

$$\left(\sum_{i=1}^{\infty} w_i\right)\left(\sum_{i=1}^{\infty} w_i a_i bi\right) \geq \left(\sum_{i=1}^{\infty} w_i a_i\right)\left(\sum_{i=1}^{\infty} w_i b_i\right). \tag{1.80}$$

Let us introduce *mean version of Lerch's transcendent*

$$\overline{\Phi^*}(s,q,w) = \frac{1-w}{w}\Phi^*(s,q,w).$$

Theorem 1.25 *The mean version of Lerch's transcendent is log-subadditive, i.e. for* $s,r>0$

$$\overline{\Phi^*}(s+r,q,w) \leq \overline{\Phi^*}(s,q,w)\overline{\Phi^*}(r,q,w). \tag{1.81}$$

Proof. We apply Chebyshev's inequality (1.80) for

$$a_i = \left(\frac{k+q}{i+q}\right)^s, \; b_i = \left(\frac{k+q}{i+q}\right)^r, \; w_i = w^i; \; i \in \mathbb{N}$$

i.e.

$$\frac{w}{1-w}\Phi^*(s+r,q,w) \leq \Phi^*(s,q,w)\Phi^*(r,q,w).$$

□

Remark 1.6 A similar version of the previous theorem can be proved for cut Lerch's transcendent $\overline{\Phi^*}_N(s,q,w)$.

1.3.4 Hybrid means

For a fixed $p_i \geq 0,\; a_i \geq 0,\; i = 1\ldots,N$ let us define linear functional on $C[a,b]$ with

$$A(f) = \sum_{i=1}^{n} p_i f(a_i),$$

where $a \leq \min a_i \leq \max a_i \leq b$. Then, the next theorem is valid.

Theorem 1.26 *For a continuous function* $g : [a,b] \to \mathbb{R}_+$, *the function* $t \mapsto A(g^t)$ *is exponentially convex on* $(0,\infty)$ *and for positive* $\alpha_1,\alpha_2,\beta_1,\beta_2$; $\alpha_1 \leq \alpha_2,\; \beta_1 \leq \beta_2,\; \alpha_1 \neq \beta_1,\; \alpha_2 \neq \beta_2$,

$$\left(\frac{A(g^{\alpha_1})}{A(g^{\alpha_2})}\right)^{\frac{1}{\alpha_1-\beta_1}} \leq \left(\frac{A(g^{\alpha_2})}{A(g^{\beta_2})}\right)^{\frac{1}{\alpha_2-\beta_2}}. \tag{1.82}$$

Proof. For a fixed $n \in \mathbb{N},\; u_i \in \mathbb{R},\; t_i > 0,\; i = 1,\ldots,n$, we define an auxiliary function

$$\Psi(x) = \sum_{i,j=1}^{n} u_i u_j x^{\frac{t_i+t_j}{2}}.$$

Since $\Psi(x) = \left(\sum\limits_{i=1}^{n} u_i x^{\frac{t_i}{2}}\right)^2 \geq 0$, we have $A(\Psi(g)) \geq 0$, i.e.

$$\sum_{i,j=1}^{n} u_i u_j A\left(g^{\frac{t_i+t_j}{2}}\right) \geq 0$$

concluding exponential convexity of the function $t \mapsto A(g^t)$. Since exponential convexity implies log-convexity (see [5]), (1.82) follows from [15] pp. 7. □

Remark 1.7 For fixed $t > 0$, let $m = \min\limits_{x\in[a,b]} g^t(x)$, $M = \max\limits_{x\in[a,b]} g^t(x)$. Then, from $A(g^t - m) \geq 0$ and $A(M - g^t) \geq 0$ it follows

$$mA(1) \leq A(g^t) \leq MA(1).$$

By the mean value theorem it follows that exists $\xi \in [a,b]$ such that

$$g^t(\xi) = \frac{A(g^t)}{A(1)}.$$

Also, for fixed $\alpha,\ \beta \in (0,\infty), \alpha \neq \beta$, following the very standard technique, we can also prove that there exists $\eta \in [a,b]$ such that

$$g^{\alpha-\beta}(\eta) = \frac{A(g^\alpha)}{A(g^\beta)}. \tag{1.83}$$

Now, if $g(\eta) \in [a,b]$, then the expression

$$\left(\frac{A(g^\alpha)}{A(g^\beta)}\right)^{\frac{1}{\alpha-\beta}}. \tag{1.84}$$

stands for the mean and, as (1.82) shows, these means have monotonicity property.

For fixed, $k \in \mathbb{N}$, $q \geq 0$, let us take $p_i = w^i$, $a_i = \frac{k+q}{i+q}$, $i = 1,\ldots,N$, $g = \mathrm{id}$. Using Remark (1.7) and (1.67), we can define hybrid means

$$H(\alpha,\beta) = \begin{cases} \left(\frac{f_h(w,N,k,q,\beta)}{f_h(w,N,k,q,\alpha)}\right)^{\frac{1}{\alpha-\beta}}, & \alpha \neq \beta; \\ \exp\left(-\frac{\frac{d}{d\alpha} f_h(w,N,k,q,\alpha)}{f_h(w,N,k,q,\alpha)}\right), & \alpha = \beta. \end{cases} \tag{1.85}$$

Theorem 1.27 *For* $0 < \alpha_1 \leq \alpha_2,\ 0 < \beta_1 \leq \beta_2,\ \alpha_1 \neq \beta_1,\ \alpha_2 \neq \beta_2$;

$$H(\alpha_1,\beta_1) \leq H(\alpha_2,\beta_2). \tag{1.86}$$

Bibliography

[1] G. Bastien, M. Rogalski, *Convexité, complète monotonie et inégalités sur les fonctions zêta et gamma, sur les fonctions de opérateurs de Baskakov et sur les fonctions arithmétiques*, Can. J. Math., 54(5) (2002), 916–944.

[2] H. Bateman, A. Erdélyi, *Higher Transcendental Functions* Vol. I, New York: McGraw-Hill, 1953.

[3] P. R. Beesack, *On inequalities complementary to Jensen's* , Canad. J. Math. 35 (1983), 324–338.

[4] S. Frontier, *Diversity and structure in aquatic ecosystems*, Mar. Biol. Ann. Rev. 23(1985), 253–312.

[5] J. Jakšetić, J. Pečarić, *Exponential Convexity method*, J. Convex Anal., 20(1) (2013), 181–197.

[6] J. Jakšetić, Đ. Pečarić, J. Pečarić, *Some classical inequalities and Zipf-Mandelbrot law*, Octogon Mathematical Magazine, vol. 25(2), October 2017, pp 11–22.

[7] J. Jakšetić, Đ. Pečarić, J. Pečarić, *Some properties of Zipf-Mandelbrot law and Hurwitz $\zeta-$function*, Math. Inequal. Appl. 21 (2) (2018) 575–584.

[8] J. Jakšetić, Đ. Pečarić, J. Pečarić, *Hybrid Zipf-Mandelbrot law*, J. Math. Inequal. (accepted)

[9] L. Horváth, Đ. Pečarić, J. Pečarić *Estimations of f- and Rényi divergences by using a cyclic refinement of the Jensen's inequality*, Bull. Malays. Math. Sci. Soc. (accepted).

[10] B. Mandelbrot, *An information theory of the statistical structure of language* Communication Theory, New York. Academic Press, 1953.

[11] D. Y. Manin, *Mandelbrot's Model for Zipf's Law: Can Mandelbrot's Model Explain Zipf's Law for Language?* , 16(3)(2009), 274–285.

[12] M. Matić, C. E. M. Pearce, J. Pečarić, *Shannons and related inequalities in information theory*, In: Rassias, TM (ed.) Survey on Classical Inequalities, pp. 127–164. Kluwer Academic, Norwell (2000).

[13] D. Mouillot, A. Lepretre, *Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity*, Environmental Monitoring and Assessment, 63(2)(2000), 279–295.

[14] J. Pečarić, *Nejednakosti*, Hrvatsko matematičko društvo, Zagreb, 1996.

[15] J. Pečarić, F. Proshan, Y. L. Tong, *Convex functions partial ordering and statistical appli- cations mathematics in Science and Engineering*, Vol. 187., Academic Press, 1992.

[16] L. Tòth, *The order of convergence of some sequences*, Mat. Lapok (Cluj) 94(1989), 133–138.

[17] M. Visser, *Zipf's law, power laws and maximum entropy*, New J. Phys 15, 043021 (2013).

[18] J. B. Wilson, *Methods for fitting dominance/diversity curves*, 2(1991), 35–46.

[19] G. K. Zipf, *Human Behavior and the Principle of least Effort*, Addison-Wesley, Reading, MA, 1949.

Chapter 2

On Zipf-Mandelbrot entropy

Muhammad Adil Khan, Đilda Pečarić, Josip Pečarić, and Zaid Mohammad Al-Sahwi

Abstract. In this paper we present several results for Shannon entropy. By using these results we give inequalities for Zipf-Mandelbrot entropy. We also discuss the results of Shannon entropy for different parametric Zipf-Mandelbrot laws. At the end we give example which shows that some of the results for Shannon entropy can not be applied for Zipf-Mandelbrot entropy.

2.1 Introduction

Zipf law is one of the necessary law in information science and is very frequently utilized in linguistics[2, 3, 20, 4, 5]. George Zipf [1] observed in the study of human language that the size of the **k**-th biggest occurrence of the event is inversely proportional to it's rank to according to

$$f(k) = \frac{c}{k^s}, \tag{2.1}$$

where c is a normalizing constant for the corpus, $f(k)$ is the number of occurrences of the k-th ranked and $s > 0$ is close to unite.

2000 Mathematics Subject Classification. 26D15, 94A17, 94A15.
Key words and phrases. Shannon entropy, Zipf-Mandelbrot entropy, convex function, Jensen's inequality.

If $n \in \{1,2,\ldots\}$, $s > 0$ and $k \in \{1,2,\ldots,n\}$, then Zipf's law (probability mass function) is defined by: $f(k,n,s) = \frac{1/k^s}{H_{n,s}}$, where $H_{n,s}$ is the n-th generalized Harmonic number [7].

The more general model introduced Benoit Mandelbrot [[8] pp. 503–512], by utilizing arguments on the fractal structure lexical trees: $g(i) = \frac{c}{(i+h)^r}$, when $h = 0$, we obtain Zipf's law.

If $n \in \mathbb{N}$, $r > 0$, $h \geq 0$ and $i \in \{1,2,\ldots,n\}$, then Zipf-Mandelbrot law (probability mass function) is defined by

$$G(i,n,h,r) = \frac{1}{(i+h)^r.H_{n,h,r}}.$$

The formula for Zipf-Mandelbrot entropy is given by

$$Z(H,h,r) = \frac{r}{H_{n,h,r}} \sum_{i=1}^{n} \frac{\log(i+h)}{(i+h)^r} + \log H_{n,h,r}, \tag{2.2}$$

where $H_{n,h,r} = \sum_{i=1}^{n} \frac{1}{(i+h)^r}$.

There are many applications of Zipf-Mandelbrot law which can be found in ecological field studies [29], and also applicable in information sciences [28]. Recently, Zipf-Mandelbrot law, Zipf-Mandelbrot entropy have been applied to various types of distances and f-divergences, for example Kullback-Leibler divergence, Bhattacharyya distance (via coefficient), Hellinger distance, x^2-divergence, Csiszâr divergence, etc, [11, 12].

2.2 Preliminary results on Shannon inequality

In this section we give some basic inequalities for Shannon entropy from [17].

A fundamental result related to the notion of the Shannon entropy is the inequality

$$\sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i}\right) \leq \sum_{i=1}^{n} p_i \log\left(\frac{1}{q_i}\right), \tag{2.3}$$

which holds for all positive real numbers p_i, q_i with

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1.$$

Throughout this chapter "log" denotes the logarithmic function taken to a fixed base $b > 1$. Equality holds in (2.3) if and only if $p_i = q_i$ for all i. For details see [[15], pp. 635–650]. The following theorem [[16], pp. 278–279] extend (2.3) and we can call it Shannon's inequality.

Theorem 2.1 *Let I be a finite or countable set of integers and $\{p_i, i \in I\}$ a set of positive real numbers such that $\sum_{i \in I} p_i = 1$. If $\{q_i, i \in I\}$ is a set of nonnegative real numbers with*

$\sum_{i\in I} q_i = \alpha > 0$, *then*

$$\sum_{i\in I} p_i \log\left(\frac{1}{p_i}\right) \le \sum_{i\in I} p_i \log\left(\frac{1}{q_i}\right) + \log\alpha, \tag{2.4}$$

with equality if and only if $q_i = \alpha p_i$ *for all* $i \in I$.

Theorem 2.2 *Let* I *be a finite or countable set of integers and* $\{p_i, i \in I\}$ *and* $\{q_i, i \in I\}$ *sets of positive numbers such that* $\sum_{i\in I} p_i = 1$ *and* $\alpha := \sum_{i\in I} q_i < \infty$. *If* $S_q = \sum_{i\in I} p_i \log\left(\frac{1}{q_i}\right)$ *is a finite, then* $S_p = \sum_{i\in I} p_i \log\left(\frac{1}{p_i}\right)$ *is also finite and*

$$0 < S_p \le S_q + \log\alpha.$$

If in addition $\sum_{i\in I} p_i^2/q_i < \infty$, *then we have*

$$\begin{aligned} 0 &\le S_q - S_p + \log\alpha \\ &\le \log\left[\alpha \sum_{i\in I} \frac{{p_i}^2}{q_i}\right] \\ &\le \frac{1}{\ln b}\left[\alpha \sum_{i\in I} \frac{{p_i}^2}{q_i} - 1\right]. \end{aligned} \tag{2.5}$$

with equality throughout if and only if $q_i = \alpha p_i$ *for all* $i \in I$.

Theorem 2.3 *Let assumptions of Theorem 2.2 be satisfied and let*

$$0 < m \le p_i/q_i \le M \quad \textit{for all} \quad i \in I.$$

Then we have

$$0 \le S_q - S_p + \log\alpha \le \log\frac{(M+m)^2}{4Mm} \le \frac{1}{4\ln b}.\frac{(M-m)^2}{Mm}. \tag{2.6}$$

Also, if $M/m \le \Phi(\varepsilon)$ *for some* $\varepsilon > 0$, *then*

$$0 \le S_q - S_p + \log\alpha \le \varepsilon.$$

Theorem 2.4 *Under the assumptions of Theorem 2.3 we have*

$$\begin{aligned} 0 &\le S_q - S_p + \log\alpha \\ &\le \log\left[\alpha(\sqrt{M} - \sqrt{m})^2 + 1\right] \\ &\le \frac{\alpha}{\ln b}(\sqrt{M} - \sqrt{m})^2. \end{aligned} \tag{2.7}$$

First we consider a discrete valued random variable X with finite range $\{x_i\}_i^r$. Assume that $p_i = P\{X = x_i\} > 0$ for $i = 1, \cdots, r$. The $b-$entropy of X is defined by

$$H_b(X) = \sum_{i=1}^{r} p_i \log\left(\frac{1}{p_i}\right), b > 1.$$

Let $L(\boldsymbol{p},\boldsymbol{q}) = \sum\limits_{i\in I} p_i \log\left(\frac{p_i}{q_i}\right)$ be represent the discrimination, Kulbackack-Leibler distance or relative entropy between the probability distribution $\boldsymbol{p} = \{p_i\}_{i\in I}$ and $\boldsymbol{q} = \{q\}_{i\in I}$. Also, let $H(\boldsymbol{p}) = S$ be the entropy and $d(\boldsymbol{p},\boldsymbol{q}) = \sum\limits_{i\in I} |p_i - q_i|$ be the variational distance between probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$.

Theorem 2.5 *With the above notations*

$$\begin{aligned} L(\boldsymbol{p},\boldsymbol{q}) + H(\boldsymbol{p}) &\geq -\log\left[1 - \frac{1}{2}d(\boldsymbol{p},\boldsymbol{q})\right] \\ &\geq \frac{1}{2\ln b} d(\boldsymbol{p},\boldsymbol{q}). \end{aligned} \tag{2.8}$$

The following bounds on the entropy function give a further improvement of Theorem 2.1.

Theorem 2.6 (a) : *Suppose X is a discrete valued random variable with finite range, we have*

$$\begin{aligned} 0 \leq \log r - H_b(X) &\leq \log\left[r \sum_{k=1}^{r} p_k^2\right] \\ &\leq \frac{1}{\ln b}\left[\sum_{k=1}^{r} p_k^2 - 1\right]. \end{aligned} \tag{2.9}$$

Equality holds throughout if and only if $p_i = \frac{1}{r}$ for $i = 1, 2, \cdots, r$.
(b) : *If $\rho = \max_{i,k} \frac{p_i}{p_k}$, then*

$$\begin{aligned} 0 \leq \log r - H_b(X) &\leq \log\left[\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right] \\ &\leq \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \end{aligned} \tag{2.10}$$

If $\rho < \Phi(\varepsilon)$ for $\varepsilon > 0$, then

$$0 \leq \log r - H_b(X) \leq \varepsilon. \tag{2.11}$$

In the following theorem $[x]$ denotes the largest integer less than or equal to x.

Theorem 2.7 *With X as above, define $M = \max_i p_i$ and $m = \min_i p_i$, then*

$$0 \leq \log_b r - H_b(X) \leq \log_b\left\{\left[\frac{r^2}{4}\right](M-m)^2 + 1\right\} \leq \frac{(M-m)^2}{\ln b}\left[\frac{r^2}{4}\right]. \tag{2.12}$$

If

$$\max_{1\le i\le j\le r} |p_i - p_j| \le \sqrt{\frac{b^{\varepsilon}-1}{\left[\frac{r^2}{4}\right]}}, \tag{2.13}$$

then (2.10) holds.

Theorem 2.8 *Let X be a discrete valued random variable with finite range $\{x_i\}_{i=1}^r$ and probability distribution $p_k = P\{X = x_k\} > 0$ for $1 \le k \le r$ and set $\rho = \max_{i,k} \frac{p_i}{p_k}$. Let δ be a permutation of $(1,2,\cdots,r)$ such that $\left(p_{\delta(k)}\right)_1^r$ is monotone. Define $P_k = \sum_{i=1}^{k} p_{\delta(i)}$ and $M = \max_{1\le k\le r} P_k(1-P_k)$. Then*

$$0 \le \log r - H_b(X) \le \log\left[M\left(\sqrt{\rho}-\frac{1}{\sqrt{\rho}}\right)^2+1\right] \le \frac{M}{\ln b}\left(\sqrt{\rho}-\frac{1}{\sqrt{\rho}}\right)^2. \tag{2.14}$$

If $\rho \le \Phi_M(\varepsilon)$ for some $\varepsilon > 0$, then

$$0 < \log r - H_b(X) \le \varepsilon. \tag{2.15}$$

In the case when X is a discrete random variable with countable range $\{x_i\}_{i\ge r}$ and probability distribution $p_i = P\{X = x_i\} > 0$ $(\sum_{i=1}^{\infty} p_i = 1)$. The b-entropy of X is defined by

$$H_b(X) = \sum_{i=1}^{\infty} p_i \log\left(\frac{1}{p_i}\right), b > 1. \tag{2.16}$$

Consider

$$\nu = \sum_{i=1}^{\infty} i p_i \tag{2.17}$$

Theorem 2.9 *Let X be a discrete valued random variable countable range $\{x_i\}_{i\ge 1}$ and probability distribution $p_i = P\{X = x_i\} > 0$ $(\sum_{i=1}^{\infty} p_i = 1)$ such that $\nu < \infty$. Then the entropy $H_b(X)$ defined by (7.48) is finite and*

$$0 < H_b(X) \le \log\frac{\nu^{\nu}}{(\nu-1)^{\nu-1}}.$$

If in addition $\sum_{i=1}^{\infty}[\nu^i/(\nu-1)^{i-1}]p_i^2 < \infty$, then

$$\begin{aligned} 0 &\le \log\frac{\nu^{\nu}}{(\nu-1)^{\nu-1}} - H_b(X) \\ &\le \log\left[\sum_{i=1}^{\infty}\frac{\nu^i}{(\nu-1)^{i-1}}p_i^2\right] \\ &\le \frac{1}{\ln b}\left[\sum_{i=1}^{\infty}\frac{\nu^i}{(\nu-1)^{i-1}}p_i^2 - 1\right], \end{aligned} \tag{2.18}$$

with equalities throughout if and only if $p_i = (\nu-1)^{i-1}/\nu^i$ for all $i \in N$.

Theorem 2.10 *(i) Under the assumptions of Theorem 2.9. If*

$$0 < L \leq \frac{v^i}{(v-1)^{i-1}} p_i \leq U \quad \text{for all} \quad i \in N,$$

then

$$\begin{aligned} 0 &\leq \log K - H_b(X) \\ &\leq \log \frac{(U+L)^2}{4LU} \\ &\leq \frac{1}{4\ln b} \frac{(U-L)^2}{LU}. \end{aligned} \tag{2.19}$$

Also, if $U/L \leq \Phi(\varepsilon)$ for some $\varepsilon > 0$, then

$$0 \leq \log K - H_b(X) \leq \varepsilon.$$

(ii) Further, we have

$$\begin{aligned} 0 &\leq \log K - H_b(X) \\ &\leq \log \left[(\sqrt{U} - \sqrt{L})^2 + 1 \right] \\ &\leq \frac{1}{\ln b} (\sqrt{U} - \sqrt{L})^2. \end{aligned} \tag{2.20}$$

Theorem 2.11 *Let X be a discrete valued random variable with finite range $R = \{x_i, i \geq I\}$ and probability distribution $p_i = P\{X = x_i\} > 0$ $(\sum_{i\in I} p_i = 1)$, β an arbitrary real number. And let that $A := \sum_{i\in I} b^{-\beta f(x_i)}$ and $E[f(x)] := \sum_{i\in I} p_i f(x_i)$ are finite. If $\sum\limits_{i\in I} p_i^2 b^{\beta f(x_i)} < \infty$. Then*

$$\begin{aligned} 0 &\leq \beta E[f(x)] - H_b(X) + \log A \\ &\leq \log \left[A \sum_{i\in I} p_i^2 b^{\beta f(x_i)} \right] \\ &\leq \frac{1}{\ln b} \left[A \sum_{i\in I} p_i^2 b^{\beta f(x_i)} - 1 \right], \end{aligned} \tag{2.21}$$

with equalities throughout if and only if $p_i = A^{-1} b^{-\beta f(x_i)}$ for all $i \in I$. Furthermore, if there are constants $L, U > 0$ such that $L \leq p_i b^{\beta f(x_i)} \leq U$ for all $i \in I$, then

$$\begin{aligned} 0 &\leq \beta E[f(x)] - H_b(X) + \log A \\ &\leq \log \frac{(U+L)^2}{4LU} \\ &\leq \frac{1}{4\ln b} \frac{(U-L)^2}{LU}. \end{aligned} \tag{2.22}$$

Also, if $U/L \le \Phi(x)$ for some $\varepsilon > 0$,

$$0 \le \beta E[f(x)] - H_b(X) + \log \le \varepsilon.$$

Further, we have

$$\begin{aligned} 0 &\le \beta E[f(x)] - H_b(X) + \log A \\ &\le \log\left[A(\sqrt{U} - \sqrt{L})^2 + 1\right] \\ &\le \frac{A}{\ln b}(\sqrt{U} - \sqrt{L})^2. \end{aligned} \tag{2.23}$$

Remark 2.1 In [17], there are serval results presented for Shannon entropy involving $\lambda = \sup_{n \in \mathbb{N}} a_n$, where

$$a_n = \left(p_n^{-1} \sum_{k=n+1}^{\infty} p_k\right), p_k > 0, \sum_{k=1}^{\infty} p_k = 1. \tag{2.24}$$

But we will show those results are not valid for Zipf-Mandelbrot law.

2.3 Inequalities for Zipf-Mandelbrot entropy

In this section we give applications of results presented in Section 2 for Zipf-Mandelbrot entropy.

Theorem 2.12 *Let $n \in \{1,2,3,\ldots\}$, $q \ge 0$, $s > 0$, $q_i > 0$, $i = 1,2,\ldots,n$ with $\sum_{i=1}^{n} q_i = 1$, then*

$$Z(H,q,s) \le \sum_{i=1}^{n} \frac{\log\left(\frac{1}{q_i}\right)}{(i+q)^s H_{n,q,s}}, \tag{2.25}$$

Proof. Replacing p_i by $\frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,\ldots,n$, we have

$$\begin{aligned} -\sum_{i=1}^{n} p_i \log p_i &= -\sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \log \frac{1}{(i+q)^s H_{n,q,s}} \\ &= \sum_{i=1}^{n} \frac{\log[((i+q)^s H_{n,q,s}]}{(i+q)^s H_{n,q,s}} \\ &= \sum_{i=1}^{n} \frac{s \log(i+q)}{(i+q)^s H_{n,q,s}} + \sum_{i=1}^{n} \frac{\log H_{n,q,s}}{(i+q)^s H_{n,q,s}} \\ &= \frac{s}{H_{n,q,s}} \sum_{i=1}^{n} \frac{\log(i+q)}{(i+q)^s} + \frac{\log H_{n,q,s}}{H_{n,q,s}} \sum_{i=1}^{n} \frac{1}{(i+q)^s}. \end{aligned}$$

Then

$$\sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i}\right) = Z(H,q,s),$$

where $H_{n,q,s} = \sum\limits_{i=1}^{n} \frac{1}{(i+q)^s}$.

Therefore, for $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$, in (2.3) we obtain (2.25). □

Theorem 2.13 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $q_i > 0$, $i = 1,2,\ldots,n$ with $\sum_{i=1}^{n} q_i = \alpha$, then*

$$Z(H,q,s) \leq \sum_{i=1}^{n} \frac{\log\left(\frac{1}{q_i}\right)}{(i+q)^s H_{n,q,s}} + \log\alpha, \tag{2.26}$$

Equality holds in (2.26) if and only if $q_i = \frac{\alpha}{(i+q)^s H_{n,q,s}}$ for all $i = 1,2,\ldots,n$.

Proof. Accordingly to the proof of Theorem 2.12, using (2.4) for $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,\ldots,n$, we get (2.26). □

Theorem 2.14 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $q_i > 0$, $i = 1,2,\ldots,n$ with $\sum_{i=1}^{n} q_i = \alpha$, then*

$$\begin{aligned}
0 &\leq \sum_{i=1}^{n} \frac{\log\left(\frac{1}{q_i}\right)}{(i+q)^s H_{n,q,s}} + \log\alpha - Z(H,q,s) \\
&\leq \log\left[\alpha \sum_{i=1}^{n} \frac{1}{q_i\left((i+q)^s H_{n,q,s}\right)^2}\right] \\
&\leq \frac{1}{\ln b}\left[\alpha \sum_{i=1}^{n} \frac{1}{q_i\left((i+q)^s H_{n,q,s}\right)^2} - 1\right],
\end{aligned} \tag{2.27}$$

with equality throughout if and only if $q_i = \frac{\alpha}{(i+q)^s H_{n,q,s}}$ for all $i = 1,2,\ldots,n$.

Proof. By taking $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$ in (2.5), $i = 1,2,\ldots,n$, we obtain we obtain the required result. □

Theorem 2.15 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $q_i > 0$, $i = 1,2,\ldots,n$ with $\sum_{i=1}^{n} q_i = \alpha$ and let*

$$0 < m \leq \frac{1}{q_i\left((i+q)^s H_{n,q,s}\right)} \leq M \quad \textit{for all} \quad i = 1,2,\ldots,n.$$

Then we have

$$0 \leq \sum_{i=1}^{n} \frac{\log\left(\frac{1}{q_i}\right)}{(i+q)^s H_{n,q,s}} + \log\alpha - Z(H,q,s) \leq \log\frac{(M+m)^2}{4Mm} \leq \frac{1}{4\ln b} \cdot \frac{(M-m)^2}{Mm}. \tag{2.28}$$

Also, if $M/m \leq \Phi(\varepsilon)$ for some $\varepsilon > 0$, then

$$0 \leq \sum_{i=1}^{n} \frac{\log\left(\frac{1}{q_i}\right)}{(i+q)^s H_{n,q,s}} + \log\alpha - Z(H,q,s) \leq \varepsilon.$$

Proof. By taking $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$ in (2.6), $i = 1,2,\ldots,n$, we get Theorem 2.15. □

Theorem 2.16 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $i = 1,2,\ldots,n$ with $\sum_{i=1}^{n} q_i = \alpha$ and let*

$$0 < m \leq \frac{1}{q_i\,((i+q)^s H_{n,q,s})} \leq M \quad \textit{for all} \quad i = 1,2,\ldots,n.$$

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} \frac{\log\left(\frac{1}{q_i}\right)}{(i+q)^s H_{n,q,s}} + \log\alpha - Z(H,q,s) \\ &\leq \log\left[\alpha(\sqrt{M}-\sqrt{m})^2 + 1\right] \\ &\leq \frac{\alpha}{\ln b}(\sqrt{M}-\sqrt{m})^2. \end{aligned} \tag{2.29}$$

Proof. By taking $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$ in (2.7), $i = 1,2,\ldots,n$, we get (2.29). □

Theorem 2.17 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $q_i > 0$, $i = 1,2,\ldots,n$ with $\sum_{i=1}^{n} q_i = \alpha$, then we have*

$$\begin{aligned} \sum_{i=1}^{n} \frac{\log[q_i((i+q)^s H_{n,q,s})]}{(i+q)^s H_{n,q,s}} - Z(H,q,s) &\leq \log\left[1 - \frac{1}{2}\sum_{i=1}^{n}\left|\frac{1-q_i[(i+q)^s H_{n,q,s}]}{(i+q)^s H_{n,q,s}}\right|\right] \\ &\leq \frac{-1}{2\ln b}\sum_{i=1}^{n}\left|\frac{1-q_i[(i+q)^s H_{n,q,s}]}{(i+q)^s H_{n,q,s}}\right|. \end{aligned} \tag{2.30}$$

OR

$$\begin{aligned} \sum_{i=1}^{n} \frac{\log q_i}{(i+q)^s H_{n,q,s}} &\leq \log\left[1 - \frac{1}{2}\sum_{i=1}^{n}\left|\frac{1-q_i[(i+q)^s H_{n,q,s}]}{(i+q)^s H_{n,q,s}}\right|\right] \\ &\leq \frac{-1}{2\ln b}\sum_{i=1}^{n}\left|\frac{1-q_i[(i+q)^s H_{n,q,s}]}{(i+q)^s H_{n,q,s}}\right|. \end{aligned} \tag{2.31}$$

Proof. By taking $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$ in (2.8), $i = 1,2,\ldots,n$, we get Theorem 2.17. □

Theorem 2.18 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $i = 1,2,\ldots,n$, then we have*

$$\begin{aligned} 0 \leq \log n - Z(H,q,s) &\leq \log\left[n\sum_{i=1}^{n}\frac{1}{[(i+q)^s H_{n,q,s}]^2}\right] \\ &\leq \frac{1}{\ln b}\left[\sum_{i=1}^{n}\frac{1}{[(i+q)^s H_{n,q,s}]^2} - 1\right]. \end{aligned} \tag{2.32}$$

Equality holds throughout if and only if $\frac{1}{(i+q)^s H_{n,q,s}} = \frac{1}{n}$ for $i = 1,2,\cdots,n$.
If $\rho = \max_{i,k} \frac{(k+q)^s H_{n,q,s}}{(i+q)^s H_{n,q,s}}$, then

$$\begin{aligned} 0 \leq \log n - Z(H,q,s) &\leq \log\left[\frac{1}{4}\left(\sqrt{\rho} + \frac{1}{\sqrt{\rho}}\right)^2\right] \\ &\leq \frac{1}{4\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \end{aligned} \tag{2.33}$$

If $\rho < \Phi(\varepsilon)$ for $\varepsilon > 0$, then

$$0 \leq \log n - Z(H,q,s) \leq \varepsilon. \tag{2.34}$$

Proof. By taking $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$ in (2.9) and (2.10), $i = 1,2,\ldots,n$, we get Theorem 2.18.
□

In the following theorem $[x]$ denotes the largest integer less than or equal to x.

Theorem 2.19 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $i = 1,2,\ldots,n$, $M = \max_i \frac{1}{(i+q)^s H_{n,q,s}}$ and $m = \min_i \frac{1}{(i+q)^s H_{n,q,s}}$, then*

$$0 \leq \log n - Z(H,q,s) \leq \log\left\{\left[\frac{n^2}{4}\right](M-m)^2 + 1\right\} \leq \frac{(M-m)^2}{\ln b}\left[\frac{n^2}{4}\right]. \tag{2.35}$$

Theorem 2.20 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $i = 1,2,\ldots,n$, $\rho = \max_{i,k} \frac{(k+q)^s H_{n,q,s}}{(i+q)^s H_{n,q,s}}$. Define $M = \max_{1\leq k\leq n} \sum_{i=1}^{k} \frac{1}{(i+q)^s H_{n,q,s}}\left(1 - \sum_{i=1}^{k}\frac{1}{(i+q)^s H_{n,q,s}}\right)$. Then*

$$0 \leq \log n - Z(H,q,s) \leq \log\left[M\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2 + 1\right] \leq \frac{M}{\ln b}\left(\sqrt{\rho} - \frac{1}{\sqrt{\rho}}\right)^2. \tag{2.36}$$

If $\rho \leq \Phi_M(\varepsilon)$ for some $\varepsilon > 0$, then

$$0 < \log n - Z(H,q,s) \leq \varepsilon. \tag{2.37}$$

Theorem 2.21 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $i = 1,2,\ldots$, and $\nu := \sum_{i=1}^{\infty} \frac{i}{(i+q)^s H_{n,q,s}} < \infty$. Then $Z(H,q,s)$ is finite and*

$$0 < Z(H,q,s) \leq \log \frac{\nu^{\nu}}{(\nu-1)^{\nu-1}}.$$

If in addition $\sum\limits_{i=1}^{\infty} \frac{\nu^i/(\nu-1)^{i-1}}{[(i+q)^s H_{n,q,s}]^2} < \infty$, then

$$\begin{aligned} 0 &\leq \log \frac{\nu^{\nu}}{(\nu-1)^{\nu-1}} - Z(H,q,s) \\ &\leq \log \left[\sum_{i=1}^{\infty} \frac{\nu^i}{[(i+q)^s H_{n,q,s}]^2 (\nu-1)^{i-1}} \right] \\ &\leq \frac{1}{\ln b} \left[\sum_{i=1}^{\infty} \frac{\nu^i}{[(i+q)^s H_{n,q,s}]^2 (\nu-1)^{i-1}} - 1 \right], \end{aligned} \tag{2.38}$$

with equalities throughout if and only if $\frac{1}{(i+q)^s H_{n,q,s}} = (\nu-1)^{i-1}/\nu^i$ for all $i \in N$.

Theorem 2.22 *(i) Under the assumptions of Theorem 2.21. If*

$$0 < L \leq \frac{\nu^i}{(i+q)^s H_{n,q,s} (\nu-1)^{i-1}} \leq U \quad \text{for all} \quad i \in N,$$

then

$$\begin{aligned} 0 &\leq \log K - Z(H,q,s) \\ &\leq \log \frac{(U+L)^2}{4LU} \\ &\leq \frac{1}{4\ln b} \frac{(U-L)^2}{LU}. \end{aligned} \tag{2.39}$$

Also, if $U/L \leq \Phi(\varepsilon)$ for some $\varepsilon > 0$, then

$$0 \leq \log K - Z(H,q,s) \leq \varepsilon.$$

(ii) Further, we have

$$\begin{aligned} 0 &\leq \log K - Z(H,q,s) \\ &\leq \log \left[(\sqrt{U} - \sqrt{L})^2 + 1 \right] \\ &\leq \frac{1}{\ln b} (\sqrt{U} - \sqrt{L})^2. \end{aligned} \tag{2.40}$$

Theorem 2.23 *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s > 0$, $i = 1,2,\ldots,n$, β an arbitrary real number. If $\sum\limits_{i=1}^{n} \frac{b^{\beta f(x_i)}}{((i+q)^s H_{n,q,s})^2} < \infty$. Then*

$$0 \leq \beta \sum_{i=1}^{n} \frac{f(x)}{(i+q)^s H_{n,q,s}} - Z(H,q,s) + \log \left(\sum_{i=1}^{n} b^{-\beta f(x_i)} \right)$$

$$
\begin{aligned}
&\leq \log\left[\sum_{i=1}^{n} b^{-\beta f(x_i)} \sum_{i=1}^{n} \frac{b^{\beta f(x_i)}}{((i+q)^s H_{n,q,s})^2}\right] \\
&\leq \frac{1}{\ln b}\left[\sum_{i=1}^{n} b^{-\beta f(x_i)} \sum_{i=1}^{n} \frac{b^{\beta f(x_i)}}{((i+q)^s H_{n,q,s})^2} - 1\right], \qquad (2.41)
\end{aligned}
$$

with equalities throughout if and only if $\frac{1}{(i+q)^s H_{n,q,s}} = \frac{b^{-\beta f(x_i)}}{\sum_{i=1}^{n} b^{-\beta f(x_i)}}$ *for all* $i = 1,2,\ldots,n$. *Furthermore, if there are constants* $L,U > 0$ *such that* $L \leq \frac{b^{\beta f(x_i)}}{(i+q)^s H_{n,q,s}} \leq U$ *for all* $i \in I$, *then*

$$
\begin{aligned}
0 &\leq \beta \sum_{i=1}^{n} \frac{f(x)}{(i+q)^s H_{n,q,s}} - Z(H,q,s) + \log\left(\sum_{i=1}^{n} b^{-\beta f(x_i)}\right) \\
&\leq \log \frac{(U+L)^2}{4LU} \\
&\leq \frac{1}{4\ln b} \frac{(U-L)^2}{LU}. \qquad (2.42)
\end{aligned}
$$

Also, if $U/L \leq \Phi(x)$ *for some* $\varepsilon > 0$,

$$
0 \leq \beta \sum_{i=1}^{n} \frac{f(x)}{(i+q)^s H_{n,q,s}} - Z(H,q,s) + \log\left(\sum_{i=1}^{n} b^{-\beta f(x_i)}\right) \leq \varepsilon.
$$

Further, we have

$$
\begin{aligned}
0 &\leq \beta \sum_{i=1}^{n} \frac{f(x)}{(i+q)^s H_{n,q,s}} - Z(H,q,s) + \log\left(\sum_{i=1}^{n} b^{-\beta f(x_i)}\right) \\
&\leq \log\left[\sum_{i=1}^{n} b^{-\beta f(x_i)} \left(\sqrt{U} - \sqrt{L}\right)^2 + 1\right] \\
&\leq \frac{\sum_{i=1}^{n} b^{-\beta f(x_i)}}{\ln b} \left(\sqrt{U} - \sqrt{L}\right)^2. \qquad (2.43)
\end{aligned}
$$

In the following theorems, we use two Zipf-Mandelbrot laws for different parameters.

Theorem 2.24 *Let* $n \in \{1,2,3,\ldots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, *then*

$$
Z(H,t_1,s_1) \leq \sum_{i=1}^{n} \frac{\log\left((i+t_2)^{s_2} H_{n,t_2,s_2}\right)}{(i+t_1)^{s_1} H_{n,t_1,s_1}}, \qquad (2.44)
$$

Proof. Let $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,\ldots,n$. Then, using the proof of Theorem 2.12, we get

$$
\sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i}\right) = \sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} = Z(H,t_1,s_1)
$$

$$\sum_{i=1}^{n} p_i \log\left(\frac{1}{q_i}\right) = \sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$$

Therefore, using (2.3) for $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,\ldots,n$, we obtain required result. □

Theorem 2.25 *Let $n \in \{1,2,3,\ldots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then*

$$\begin{aligned}
0 &\leq \sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - Z(H,t_1,s_1) \\
&\leq \log\left[\alpha \sum_{i\in I} \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{((i+t_1)^{s_1} H_{n,t_1,s_1})^2}\right] \\
&\leq \frac{1}{\ln b}\left[\alpha \sum_{i\in I} \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{((i+t_1)^{s_1} H_{n,t_1,s_1})^2} - 1\right]. \qquad (2.45)
\end{aligned}$$

Proof. Taking $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, in (2.5), $i = 1,2,\ldots,n$, we obtain (2.45). □

Theorem 2.26 *Let $n \in \{1,2,3,\ldots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then*

$$0 < m \leq \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \leq M \quad \textit{for all} \quad i = 1,2,\ldots,n.$$

Then we have

$$0 \leq \sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - Z(H,t_1,s_1) \leq \log\frac{(M+m)^2}{4Mm} \leq \frac{1}{4\ln b}.\frac{(M-m)^2}{Mm}. \qquad (2.46)$$

Also, if $M/m \leq \Phi(\varepsilon)$ for some $\varepsilon > 0$, then

$$0 \leq \sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - Z(H,t_1,s_1) \leq \varepsilon.$$

Theorem 2.27 *Let $n \in \{1,2,3,\ldots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then*

$$0 < m \leq \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \leq M \quad \textit{for all} \quad i = 1,2,\ldots,n.$$

$$\begin{aligned}
0 &\leq \sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - Z(H,t_1,s_1) \\
&\leq \log\left[\alpha(\sqrt{M}-\sqrt{m})^2 + 1\right] \\
&\leq \frac{\alpha}{\ln b}(\sqrt{M}-\sqrt{m})^2. \qquad (2.47)
\end{aligned}$$

Theorem 2.28 *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then*

$$\sum_{i=1}^{n} \frac{\log\left(\frac{(i+t_2)^{s_2}H_{n,t_2,s_2}}{(i+t_1)^{s_1}H_{n,t_1,s_1}}\right)}{(i+t_1)^{s_1}H_{n,t_1,s_1}} - Z(H,t_1,s_1) \leq \log\left[1-\frac{1}{2}\sum_{i=1}^{n}\left|\frac{(i+t_2)^{s_2}H_{n,t_2,s_2}-(i+t_1)^{s_1}H_{n,t_1,s_1}}{((i+t_2)^{s_2}H_{n,t_2,s_2})((i+t_1)^{s_1}H_{n,t_1,s_1})}\right|\right]$$
$$\leq \frac{-1}{2\ln b}\sum_{i=1}^{n}\left|\frac{(i+t_2)^{s_2}H_{n,t_2,s_2}-(i+t_1)^{s_1}H_{n,t_1,s_1}}{((i+t_2)^{s_2}H_{n,t_2,s_2})((i+t_1)^{s_1}H_{n,t_1,s_1})}\right|. \tag{2.48}$$

OR

$$\sum_{i=1}^{n} \frac{\log\left(\frac{1}{(i+t_2)^{s_2}H_{n,t_2,s_2}}\right)}{(i+t_1)^{s_1}H_{n,t_1,s_1}} \leq \log\left[1-\frac{1}{2}\sum_{i=1}^{n}\left|\frac{(i+t_2)^{s_2}H_{n,t_2,s_2}-(i+t_1)^{s_1}H_{n,t_1,s_1}}{((i+t_2)^{s_2}H_{n,t_2,s_2})((i+t_1)^{s_1}H_{n,t_1,s_1})}\right|\right]$$
$$\leq \frac{-1}{2\ln b}\sum_{i=1}^{n}\left|\frac{(i+t_2)^{s_2}H_{n,t_2,s_2}-(i+t_1)^{s_1}H_{n,t_1,s_1}}{((i+t_2)^{s_2}H_{n,t_2,s_2})((i+t_1)^{s_1}H_{n,t_1,s_1})}\right|. \tag{2.49}$$

Example 2.1 Let

$$a_n = (n+q)^s \sum_{i=n+1}^{\infty} \frac{1}{(i+q)^s} \text{ and } \lambda = \sup_{n\in N} a_n \tag{2.50}$$

Since for $i > n$ we have $\frac{q}{n} > \frac{q}{i} \Rightarrow (1+\frac{q}{n})^s > (1+\frac{q}{i})^s \Rightarrow (\frac{n+q}{n})^s > (\frac{i+q}{i})^s \Rightarrow (\frac{n+q}{i+q})^s > (\frac{n}{i})^s$. Therefore

$$a_n = (n+q)^s \sum_{i=n+1}^{\infty} \frac{1}{(i+q)^s} > n^s \sum_{i=n+1}^{\infty} \frac{1}{i^s} \tag{2.51}$$

Now for $n = 2^k$ we have

$$\begin{aligned} a_{2^k} &> (2^k)^s\left[\frac{1}{(2^k+1)^s}+\cdots+\frac{1}{(2^{k+1})^s}+\frac{1}{(2^{k+1}+1)^s}+\cdots+\frac{1}{(2^{k+2})^s}+\dots\right] \\ &\geq (2^k)^s\left[\frac{2^k}{(2^{k+1})^s}+\frac{2^{k+1}}{(2^{k+2})^s}+\dots\right] \\ &= \frac{2^k}{2^s-2}. \end{aligned}$$

If $k \to \infty$ then $\frac{2^k}{2^s-2} \to \infty$. Therefore $\lambda = \infty$.

Remark 2.2 It is obvious that the sequence a_n in Example 2.1 is the related sequence of a_n in (2.24) for Zipf Mandelbrot law. But from Example 2.1 it is clear that we can not apply results of Shannon entropy involving λ for Zipf Mandelbrot law.

Bibliography

[1] G. K. Zipf, *Human behavior and the principle of least effort: An introduction to human ecology*, Ravenio Books, (2016).

[2] G. K. Zipf, *Selected studies of the principle of relative frequency in language*, (1932).

[3] H. Kucera, W. N. Francis, *Computational analysis of present-day American English*, Dartmouth Publishing Group, (1967).

[4] L. Hřebček, *Zipf's law and text*, Glottometrics. **3**, (2002), 27–38.

[5] S. T. Piantadosi, *Zipf's word frequency law in natural language: A critical review and future directions*, Psychonomic bulletin and review. **21**, 5 (2014), 1112–1130.

[6] S. Yu, C. Xu, H. Liu, *Zipf's law in 50 languages: its structural pattern, linguistic interpretation, and cognitive motivation*, arXiv preprint arXiv:1807.01855, (2018).

[7] J. Havil, *Gamma: Exploring Euler's Constant*, (Princeton, NJ: Princeton University Press), **105**, (2003).

[8] B. Mandelbrot, *An informational theory of the statistical structure of language*, Communication theory, (1953), 84, 486–502.

[9] D. Mouillot, A. Lepretre, *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 (2000), 279–295.

[10] Z. K. Silagadze, *Citations and the Zipf-Mandelbrots law*, Complex Syst. *11*, (1997), 487–499.

[11] N. Lovričević, D. Pečarić, J. Pečarić, *Zipf-Mandelbrot law, f-divergences and the Jensen-type interpolating inequalities*, J. Inequal. Appl. 36, (2018).

[12] M. Adil Khan, D. Pečarić, J. Pečarić, *On Zipf-Mandelbrot entropy*, J. Comput. Appl. Math. **346**, (2019), 192–204.

[13] J. Jakšetić, Đ. Pečarić, J. Pečarić, *Hybrid Zipf-mandelbrot law, submitted.*

[14] M. Visser, Zipf's law, *power laws and maximum entropy*, New J. Phys **15**, 043021 (2013).

[15] D. S. Mitrinović, J. E. Pečarić, A. M. Fink, *Classical and New Inequalities in Analysis*, (1993), Mathematics and its Applications (East European Series).

[16] R. J. McEliece, *The Theory of Information and Coding*, (Encyclopedia of Mathematics and its Applications No. 86), (1977).

[17] M. Matic, C. E. Pearce, J. E. Pečarić, *Shannons and related inequalities in information theory*, in Survey on Classical Inequalities (pp. 127–164), Springer, Dordrecht, (2000).

Chapter 3

Approximating f-divergence via Hermite interpolating polynomial

Đilda Pečarić, Josip Pečarić and Ana Vukelić

Abstract. In this paper we introduce a new functional based on the f-divergence functional, and then we obtain some estimates for two special cases. We use the Cauchy's error representation of Hermite interpolating polynomial and the results concerning to the Hermite-Hadamard inequalities are presented. Zipf-Mandelbrot law is used to illustrate the results.

3.1 Introduction

We follow here notations and terminology about **Hermite interpolating polynomial** from [1, p. 62]:

Let $-\infty < a < b < \infty$, and $a \leq a_1 < a_2 < \ldots < a_r \leq b$, $(r \geq 2)$ be given. For $f \in C^n[a,b]$ a unique polynomial $P_H(t)$ of degree $(n-1)$, exists, fulfilling one of the following

2000 Mathematics Subject Classification. Primary 26D15, Secondary 94A17.
Key words and phrases. Hermite interpolating polynomial, Hermite-Hadamard inequality, Csiszár f-divergence, Kullback-Leibler divergence, Shannon entropy, Zipf-Mandelbrot law.

conditions:

Hermite conditions

$$P_H^{(i)}(a_j) = f^{(i)}(a_j);\ 0 \leq i \leq k_j,\ 1 \leq j \leq r,\ \sum_{j=1}^{r} k_j + r = n,$$

in particular:

Simple Hermite or Osculatory conditions
($n = 2m$, $r = m$, $k_j = 1$ for all j)

$$P_O(a_j) = f(a_j),\ P_O'(a_j) = f'(a_j),\ 1 \leq j \leq m,$$

Lagrange conditions ($r = n$, $k_j = 0$ for all j)

$$P_L(a_j) = f(a_j),\ 1 \leq j \leq n,$$

Type $(m, n-m)$ **conditions** ($r = 2$, $1 \leq m \leq n-1$, $k_1 = m-1$, $k_2 = n-m-1$)

$$P_{mn}^{(i)}(a) = f^{(i)}(a),\ 0 \leq i \leq m-1,$$
$$P_{mn}^{(i)}(b) = f^{(i)}(b),\ 0 \leq i \leq n-m-1,$$

Two-point Taylor conditions ($n = 2m$, $r = 2$, $k_1 = k_2 = m-1$)

$$P_{2T}^{(i)}(a) = f^{(i)}(a),\ P_{2T}^{(i)}(b) = f^{(i)}(b),\ 0 \leq i \leq m-1.$$

Divergences between probability distributions have introduced to measure the difference between them. A lot of different type of divergences exist. The following notion was introduced by Csiszár in [3] and [4]:

Definition 3.1 *Let $f : (0,\infty) \to (0,\infty)$ be a convex function, and let $\mathbf{p} := (p_1, \ldots, p_s)$ and $\mathbf{q} := (q_1, \ldots, q_s)$ be positive probability distributions. The f-divergence functional is*

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

It is possible to use nonnegative probability distributions in the f-divergence functional, by define

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

Based on the previous definition we introduce a new functional:

Definition 3.2 *Let $J \subset (0,\infty)$ be an interval, and let $f : J \to \mathbb{R}$ be a function. Let $\mathbf{p} := (p_1, \ldots, p_s) \in (0,\infty)^s$, and $\mathbf{q} := (q_1, \ldots, q_s) \in (0,\infty)^s$ such that*

$$\frac{p_i}{q_i} \in J,\ \ i = 1, \ldots, s.$$

Then let

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

We mention two special cases of the previous functional.

The first case corresponds to the entropy of a discrete probability distribution ($f :=$ ln, $\mathbf{p} := \mathbf{e} = (1,\dots,1)$):

Definition 3.3 *The Shannon entropy of a positive probability distributions* $\mathbf{p} := (p_1,\dots,p_s)$ *is defined by*

$$H(\mathbf{p}) := -\sum_{i=1,}^{s} p_i \ln p_i.$$

The second case corresponds to the relative entropy or Kullback-Leibler divergence between two probability distribution ($f := id \ln$):

Definition 3.4 *The Kullback-Leibler between the positive probability distributions* $\mathbf{p} := (p_1,\dots,p_s)$ *and* $\mathbf{q} := (q_1,\dots,q_s)$ *is defined by*

$$D(\mathbf{p}||\mathbf{q}) := \sum_{i=1}^{s} p_i \ln\left(\frac{p_i}{q_i}\right).$$

In this paper we obtain some estimations of above functionals by using the Cauchy's error representation of Hermite interpolating polynomial. As a special case, Hermite-Hadamard type inequalities, will be considered. By using Zipf-Manedelbrot law we will give the applications of these results.

For some results related to f-divergence see the papers [2] and [6].

3.2 Cauchy's error representation and inequalities for f-divergence

In [1, p. 71] the following theorem is proved:

Theorem 3.1 *Let* $F(t) \in C^{n-1}([a,b])$ *and suppose that* $F^{(n)}(t)$ *exists at each point of* (a,b). *Then*

$$F(t) - \sum_{j=1}^{r}\sum_{i=0}^{k_j} H_{ij}(t)F^{(i)}(a_j) = \frac{1}{n!}\omega(t)F^{(n)}(\xi), \tag{3.1}$$

where $\xi \in (a,b)$ *and* H_{ij} *are fundamental polynomials of the Hermite basis defined by*

$$H_{ij}(t) = \frac{1}{i!}\frac{\omega(t)}{(t-a_j)^{k_j+1-i}}\sum_{k=0}^{k_j-i}\frac{1}{k_j!}\left[\frac{(t-a_j)^{k_j+1}}{\omega(t)}\right]^{(k)}_{t=a_j}(t-a_j)^k, \tag{3.2}$$

where

$$\omega(t) = \prod_{j=1}^{r}(t-a_j)^{k_j+1}. \tag{3.3}$$

Now, using Theorem 3.1 for $\tilde{I}_{\ln}(\mathbf{p},\mathbf{q})$ and $D(\mathbf{p}||\mathbf{q})$ we get the following corollaries:

Corollary 3.1 *Let n is even, and H_{ij} are defined on $[a,b]\subseteq(0,\infty)$ by (3.2), such that k_j is odd for all $j=1,\ldots,r$. Let $\mathbf{p}:=(p_1,\ldots,p_s)\in(0,\infty)^s$, and $\mathbf{q}:=(q_1,\ldots,q_s)\in(0,\infty)^s$ such that*

$$\frac{p_i}{q_i}\in[a,b],\ \ i=1,\ldots,s.$$

Then we have

$$\tilde{I}_{\ln}(\mathbf{p},\mathbf{q})-\sum_{j=1}^{r}\left[\ln a_j\tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q})+\sum_{i=1}^{k_j}\frac{(-1)^{i-1}(i-1)!}{a_j^i}\tilde{I}_{H_{ij}}(\mathbf{p},\mathbf{q})\right]\leq 0 \tag{3.4}$$

and

$$D(\mathbf{p}||\mathbf{q})-\sum_{j=1}^{r}\Bigg[a_j\ln a_j\tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q})+(\ln a_j+1)\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q})$$
$$+\sum_{i=2}^{k_j}\frac{(-1)^{i-2}(i-2)!}{a_j^{i-1}}\tilde{I}_{H_{ij}}(\mathbf{p},\mathbf{q})\Bigg]\geq 0. \tag{3.5}$$

For n odd the inequalities are reversed.

Proof. We apply Theorem 3.1 with $J:=[a,b], F:=\ln$ for first inequality and $F:=id\ln$ for second inequality. Since k_j is odd for all $j=1,\ldots,r$, then using (3.3), we get that $\omega(t)\geq 0$. By using (3.1) for n even, (3.4) and (3.5) obviously hold. $\square$

Remark 3.1 If we put that $n=2m,\ r=m$ and $k_j=1$ for all j we get Hermite interpolating polynomial with simple Hermite or Osculatory conditions and then

$$\tilde{I}_{\ln}(\mathbf{p},\mathbf{q})-\sum_{j=1}^{m}\left[\ln a_j\tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q})+\frac{1}{a_j}\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q})\right]\leq 0$$

and

$$D(\mathbf{p}||\mathbf{q})-\sum_{j=1}^{m}\left[a_j\ln a_j\tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q})+(\ln a_j+1)\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q})\right]\geq 0.$$

Corollary 3.2 *Let n is even, and H_{ij} are defined on $[a,b]\subseteq(0,\infty)$ by (3.2), such that $a_1=a$ and k_j is odd for all $j=2,\ldots,r$. Let $\mathbf{p}:=(p_1,\ldots,p_s)\in(0,\infty)^s$, and $\mathbf{q}:=(q_1,\ldots,q_s)\in(0,\infty)^s$ such that*

$$\frac{p_i}{q_i}\in[a,b],\ \ i=1,\ldots,s.$$

Then we have

$$\tilde{I}_{\ln}(\mathbf{p},\mathbf{q})-\ln a\tilde{I}_{H_{01}}(\mathbf{p},\mathbf{q})-\sum_{i=1}^{k_1}\frac{(-1)^{i-1}(i-1)!}{a^i}\tilde{I}_{H_{i1}}(\mathbf{p},\mathbf{q})$$

$$- \sum_{j=2}^{r} \left[\ln a_j \tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + \sum_{i=1}^{k_j} \frac{(-1)^{i-1}(i-1)!}{a_j^i} \tilde{I}_{H_{ij}}(\mathbf{p},\mathbf{q}) \right] \leq 0 \tag{3.6}$$

and

$$\begin{aligned} & D(\mathbf{p}||\mathbf{q}) - a\ln a \tilde{I}_{H_{01}}(\mathbf{p},\mathbf{q}) - (\ln a+1)\tilde{I}_{H_{11}}(\mathbf{p},\mathbf{q}) \\ & - \sum_{i=2}^{k_1} \frac{(-1)^{i-2}(i-2)!}{a^{i-1}} \tilde{I}_{H_{i1}}(\mathbf{p},\mathbf{q}) - \sum_{j=2}^{r} \Big[a_j \ln a_j \tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) \\ & + (\ln a_j + 1)\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q}) + \sum_{i=2}^{k_j} \frac{(-1)^{i-2}(i-2)!}{a_j^{i-1}} \tilde{I}_{H_{ij}}(\mathbf{p},\mathbf{q}) \Big] \geq 0. \end{aligned} \tag{3.7}$$

For n odd the inequalities are reversed.

Proof. Now $\omega(t) = (t-a)^{k_1+1}\prod_{j=2}^{r}(t-a_j)^{k_j+1}$. Since k_j is odd for all $j = 2,\ldots,r$, we get that $\omega(t) \geq 0$. So, by using (3.1) for n even, (3.6) and (3.7) obviously hold. □

Corollary 3.3 *Let n is even, and H_{ij} are defined on $[a,b] \subseteq (0,\infty)$ by (3.2), such that $a_r = b$. Let $\mathbf{p} := (p_1,\ldots,p_s) \in (0,\infty)^s$, and $\mathbf{q} := (q_1,\ldots,q_s) \in (0,\infty)^s$ such that*

$$\frac{p_i}{q_i} \in [a,b], \quad i = 1,\ldots,s.$$

Then

(a) If k_j is odd for all $j = 1,\ldots,r$, we have

$$\begin{aligned} & \tilde{I}_{\ln}(\mathbf{p},\mathbf{q}) - \sum_{j=1}^{r-1} \left[\ln a_j \tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + \sum_{i=1}^{k_j} \frac{(-1)^{i-1}(i-1)!}{a_j^i} \tilde{I}_{H_{ij}}(\mathbf{p},\mathbf{q}) \right] \\ & - \ln b \tilde{I}_{H_{0r}}(\mathbf{p},\mathbf{q}) - \sum_{i=1}^{k_r} \frac{(-1)^{i-1}(i-1)!}{b^i} \tilde{I}_{H_{ir}}(\mathbf{p},\mathbf{q}) \leq 0 \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} & D(\mathbf{p}||\mathbf{q}) - \sum_{j=1}^{r-1} \left[a_j \ln a_j \tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + (\ln a_j+1)\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q}) + \sum_{i=2}^{k_j} \frac{(-1)^{i-2}(i-2)!}{a_j^{i-1}} \tilde{I}_{H_{ij}}(\mathbf{p},\mathbf{q}) \right] \\ & - b\ln b \tilde{I}_{H_{0r}}(\mathbf{p},\mathbf{q}) - (\ln b+1)\tilde{I}_{H_{1r}}(\mathbf{p},\mathbf{q}) - \sum_{i=2}^{k_r} \frac{(-1)^{i-2}(i-2)!}{b^{i-1}} \tilde{I}_{H_{ir}}(\mathbf{p},\mathbf{q}) \geq 0. \end{aligned} \tag{3.9}$$

(b) If k_j is odd for all $j = 1,\ldots,r-1$ and k_r is even, we have the reversed inequalities. For n odd the inequalities are reversed.

Proof. Now $\omega(t) = (t-b)^{k_r+1}\prod_{j=1}^{r-1}(t-a_j)^{k_j+1}$.

(a) Since k_j is odd for all $j = 1,\ldots,r$, we get that $\omega(t) \geq 0$.

(b) Since k_j is odd for all $j = 1,\ldots,r-1$ and k_r is even, we get that $\omega(t) \leq 0$.

So, by using (3.1) for n even, (3.8) and (3.9) obviously hold. □

Corollary 3.4 *Let n is even, and H_{ij} are defined on $[a,b] \subseteq (0,\infty)$ by (3.2), such that $a_1 = a$ and $a_r = b$. Let $\mathbf{p} := (p_1,\ldots,p_s) \in (0,\infty)^s$, and $\mathbf{q} := (q_1,\ldots,q_s) \in (0,\infty)^s$ such that*

$$\frac{p_i}{q_i} \in [a,b], \ \ i = 1,\ldots,s.$$

Then

(a) If k_j is odd for all $j = 2,\ldots,r$, we have

$$\begin{aligned} &\tilde{I}_{\ln}(\mathbf{p},\mathbf{q}) - \ln a \tilde{I}_{H_{01}}(\mathbf{p},\mathbf{q}) - \sum_{i=1}^{k_1} \frac{(-1)^{i-1}(i-1)!}{a^i} \tilde{I}_{H_{i1}}(\mathbf{p},\mathbf{q}) \\ &- \sum_{j=2}^{r-1} \left[\ln a_j \tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + \sum_{i=1}^{k_j} \frac{(-1)^{i-1}(i-1)!}{a_j^i} \tilde{I}_{H_{ij}}(\mathbf{p},\mathbf{q}) \right] \\ &- \ln b \tilde{I}_{H_{0r}}(\mathbf{p},\mathbf{q}) - \sum_{i=1}^{k_r} \frac{(-1)^{i-1}(i-1)!}{b^i} \tilde{I}_{H_{ir}}(\mathbf{p},\mathbf{q}) \le 0 \end{aligned} \tag{3.10}$$

and

$$\begin{aligned} &D(\mathbf{p}||\mathbf{q}) - a \ln a \tilde{I}_{H_{01}}(\mathbf{p},\mathbf{q}) - (\ln a + 1) \tilde{I}_{H_{11}}(\mathbf{p},\mathbf{q}) - \sum_{i=2}^{k_1} \frac{(-1)^{i-2}(i-2)!}{a^{i-1}} \tilde{I}_{H_{i1}}(\mathbf{p},\mathbf{q}) \\ &- \sum_{j=2}^{r-1} \left[a_j \ln a_j \tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + (\ln a_j + 1) \tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q}) + \sum_{i=2}^{k_j} \frac{(-1)^{i-2}(i-2)!}{a_j^{i-1}} \tilde{I}_{H_{ij}}(\mathbf{p},\mathbf{q}) \right] \\ &- b \ln b \tilde{I}_{H_{0r}}(\mathbf{p},\mathbf{q}) - (\ln b + 1) \tilde{I}_{H_{1r}}(\mathbf{p},\mathbf{q}) - \sum_{i=2}^{k_r} \frac{(-1)^{i-2}(i-2)!}{b^{i-2}} \tilde{I}_{H_{ir}}(\mathbf{p},\mathbf{q}) \ge 0. \end{aligned} \tag{3.11}$$

(b) If k_j is odd for all $j = 2,\ldots,r-1$ and k_r is even, we have the reversed inequalities. For n odd the inequalities are reversed.

Proof. Now $\omega(t) = (t-a)^{k_1+1}(t-b)^{k_r+1} \prod_{j=2}^{r-1}(t-a_j)^{k_j+1}$.

(a) Since k_j is odd for all $j = 2,\ldots,r$, we get that $\omega(t) \ge 0$.

(b) Since k_j is odd for all $j = 2,\ldots,r-1$ and k_r is even, we get that $\omega(t) \le 0$.

So, by using (3.1) for n even, (3.10) and (3.11) obviously hold. □

Remark 3.2 If we put $r = 2$, $1 \le m \le n-1$, $k_1 = m-1$, $k_2 = n-m-1$ and k_2 is even then we get Hermite interpolating polynomial with $(m, n-m)$ type conditions and then

$$\begin{aligned} &\tilde{I}_{\ln}(\mathbf{p},\mathbf{q}) - \ln a \tilde{I}_{H_{01}}(\mathbf{p},\mathbf{q}) - \sum_{i=1}^{m-1} \frac{(-1)^{i-1}(i-1)!}{a^i} \tilde{I}_{H_{i1}}(\mathbf{p},\mathbf{q}) \\ &- \ln b \tilde{I}_{H_{02}}(\mathbf{p},\mathbf{q}) - \sum_{i=1}^{n-m-1} \frac{(-1)^{i-1}(i-1)!}{b^i} \tilde{I}_{H_{i2}}(\mathbf{p},\mathbf{q}) \le 0 \end{aligned}$$

and

$$D(\mathbf{p}||\mathbf{q}) - a \ln a \tilde{I}_{H_{01}}(\mathbf{p},\mathbf{q}) - (\ln a + 1) \tilde{I}_{H_{11}}(\mathbf{p},\mathbf{q}) - \sum_{i=2}^{m-1} \frac{(-1)^{i-2}(i-2)!}{a^{i-1}} \tilde{I}_{H_{i1}}(\mathbf{p},\mathbf{q})$$

$$- \; b\ln b\tilde{I}_{H_{02}}(\mathbf{p},\mathbf{q}) - (\ln b+1)\tilde{I}_{H_{12}}(\mathbf{p},\mathbf{q}) - \sum_{i=2}^{n-m-1} \frac{(-1)^{i-2}(i-2)!}{b^{i-1}}\tilde{I}_{H_{i2}}(\mathbf{p},\mathbf{q}) \geq 0.$$

For k_2 odd, the above inequalities are reversed.

If we put $n = 2m,\ r = 2,\ k_1 = k_2 = m-1$ and m is even then we get Hermite interpolating polynomial with two-point Taylor conditions and then

$$\tilde{I}_{\ln}(\mathbf{p},\mathbf{q}) - \ln a\tilde{I}_{H_{01}}(\mathbf{p},\mathbf{q}) - \sum_{i=1}^{m-1} \frac{(-1)^{i-1}(i-1)!}{a^{i}}\tilde{I}_{H_{i1}}(\mathbf{p},\mathbf{q})$$
$$- \; \ln b\tilde{I}_{H_{02}}(\mathbf{p},\mathbf{q}) - \sum_{i=1}^{m-1} \frac{(-1)^{i-1}(i-1)!}{b^{i}}\tilde{I}_{H_{i2}}(\mathbf{p},\mathbf{q}) \leq 0$$

and

$$D(\mathbf{p}||\mathbf{q}) - a\ln a\tilde{I}_{H_{01}}(\mathbf{p},\mathbf{q}) - (\ln a+1)\tilde{I}_{H_{11}}(\mathbf{p},\mathbf{q}) - \sum_{i=2}^{m-1} \frac{(-1)^{i-2}(i-2)!}{a^{i-1}}\tilde{I}_{H_{i1}}(\mathbf{p},\mathbf{q})$$
$$- \; \ln b\tilde{I}_{H_{02}}(\mathbf{p},\mathbf{q}) - (\ln b+1)\tilde{I}_{H_{12}}(\mathbf{p},\mathbf{q}) - \sum_{i=2}^{m-1} \frac{(-1)^{i-2}(i-2)!}{b^{i-1}}\tilde{I}_{H_{i2}}(\mathbf{p},\mathbf{q}) \geq 0.$$

For m odd, the above inequalities are reversed.

Remark 3.3 By using $\mathbf{p} := \mathbf{e} = (1,\ldots,1)$ in Corollaries 3.1-3.4, $\tilde{I}_{\ln}(\mathbf{p},\mathbf{q}) = H(\mathbf{q})$ and $\tilde{I}_f(\mathbf{p},\mathbf{q}) = \tilde{I}_f(\mathbf{e},\mathbf{q})$ we get inequalities for Shannon entropy.

Also, we can notice that $-\tilde{I}_{\ln}(\mathbf{p},\mathbf{q}) = \tilde{I}_{-\ln}(\mathbf{p},\mathbf{q}) = D(\mathbf{q}||\mathbf{p})$.

3.3 Hermite-Hadamard type inequalities

As a consequences of our results given in Section 2, here we give the Hermite-Hadamard type inequalities for Csiszár f-divergence and Shannon entropy.

Let H_{ij} and $\bar{H}_{ij}$ are defined on $[a,b]$ by

$$H_{ij}(t) = \frac{1}{i!}\frac{\omega(t)}{(t-a_j)^{k_j+1-i}}\sum_{k=0}^{k_j-i}\frac{1}{k_j!}\left[\frac{(t-a_j)^{k_j+1}}{\omega(t)}\right]^{(k)}_{t=a_j}(t-a_j)^k, \tag{3.12}$$

and

$$\bar{H}_{ij}(t) = \frac{1}{i!}\frac{\bar{\omega}(t)}{(t-b_j)^{l_j+1-i}}\sum_{k=0}^{l_j-i}\frac{1}{l_j!}\left[\frac{(t-b_j)^{l_j+1}}{\bar{\omega}(t)}\right]^{(k)}_{t=b_j}(t-b_j)^k, \tag{3.13}$$

where

$$\omega(t) = \prod_{j=1}^{r}(t-a_j)^{k_j+1}, \quad \bar{\omega}(t) = \prod_{j=1}^{\bar{r}}(t-b_j)^{l_j+1} \tag{3.14}$$

for $a \leq a_1 < a_2 \leq \ldots < a_r \leq b$, $a \leq b_1 < b_2 < \ldots < b_{\bar{r}} \leq b, (r, \bar{r} \geq 2)$ and $\sum_{j=1}^{r} k_j + r = \sum_{j=1}^{\bar{r}} l_j + \bar{r} = n$.

Theorem 3.2 *Let $a_1 = a$, $b_{\bar{r}} = b$, $k_1 = 0$ and $k_j = 1$ for all $j = 2, \ldots, r$, $l_j = 1$ for all $j = 1, \ldots, \bar{r} - 1$ and $l_{\bar{r}} = 0$. Then, we have*

$$\begin{aligned}
&\ln a \tilde{I}_{H_{01}}(\mathbf{p}, \mathbf{q}) + \sum_{j=2}^{r} \left[\ln a_j \tilde{I}_{H_{0j}}(\mathbf{p}, \mathbf{q}) + \frac{1}{a_j} \tilde{I}_{H_{1j}}(\mathbf{p}, \mathbf{q}) \right] \\
&\leq \tilde{I}_{\ln}(\mathbf{p}, \mathbf{q}) \qquad (3.15) \\
&\leq \sum_{j=1}^{\bar{r}-1} \left[\ln b_j \tilde{I}_{\bar{H}_{0j}}(\mathbf{p}, \mathbf{q}) + \frac{1}{b_j} \tilde{I}_{\bar{H}_{1j}}(\mathbf{p}, \mathbf{q}) \right] + \ln b \tilde{I}_{\bar{H}_{0r}}(\mathbf{p}, \mathbf{q})
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{j=1}^{\bar{r}-1} \left[b_j \ln b_j \tilde{I}_{\bar{H}_{0j}}(\mathbf{p}, \mathbf{q}) + (\ln b_j + 1) \tilde{I}_{\bar{H}_{1j}}(\mathbf{p}, \mathbf{q}) \right] + b \ln b \tilde{I}_{\bar{H}_{0r}}(\mathbf{p}, \mathbf{q}) \\
&\leq D(\mathbf{p} || \mathbf{q}) \qquad (3.16) \\
&\leq a \ln a \tilde{I}_{H_{01}}(\mathbf{p}, \mathbf{q}) + \sum_{j=2}^{r} \left[a_j \ln a_j \tilde{I}_{H_{0j}}(\mathbf{p}, \mathbf{q}) + (\ln a_j + 1) \tilde{I}_{H_{1j}}(\mathbf{p}, \mathbf{q}) \right],
\end{aligned}$$

where

$$H_{01}(t) = \frac{P_{r-1}^2(t)}{P_{r-1}^2(a)},$$

$$H_{0j}(t) = \frac{(t-a) P_{r-1}^2(t)}{(t-a_j)^2 \left[P'_{r-1}(a_j) \right]^2 (a_j - a)} \left(1 - \frac{P'_{r-1}(a_j) + (a_j - a) P''_{r-1}(a_j)}{(a_j - a) P'_{r-1}(a_j)} (t - a_j) \right),$$

$$H_{1j}(t) = \frac{(t-a) P_{r-1}^2(t)}{(t-a_j)(a_j - a) \left[P'_{r-1}(a_j) \right]^2},$$

$$\bar{H}_{0j}(t) = \frac{(b-t) \bar{P}_{r-1}^2(t)}{(t-b_j)^2 \left[\bar{P}'_{r-1}(b_j) \right]^2 (b - b_j)} \left(1 + \frac{\bar{P}'_{r-1}(b_j) - (b - b_j) \bar{P}''_{r-1}(b_j)}{(b - b_j) \bar{P}'_{r-1}(b_j)} (t - b_j) \right),$$

$$\bar{H}_{1j}(t) = \frac{(b-t) \bar{P}_{r-1}^2(t)}{(t-b_j)(b - b_j) \left[\bar{P}'_{r-1}(b_j) \right]^2},$$

$$\bar{H}_{0r}(t) = \frac{\bar{P}_{r-1}^2(t)}{\bar{P}_{r-1}^2(b)},$$

and

$$P_{r-1}(t) = \prod_{j=2}^{r} (t - a_j), \quad \bar{P}_{r-1}(t) = \prod_{j=1}^{r-1} (t - b_j)$$

for $a < a_2 < \ldots < a_r \leq b$, $a \leq b_1 < b_2 < \ldots < b_{r-1} < b, (r \geq 2)$.

Proof. We use Corollary 3.2 and Corollary 3.3(b) for $n = 2r-1$ and then calculate

$$H_{01}(t) = \frac{(t-a)P_{r-1}^2(t)}{(t-a)} \cdot \left[\frac{(t-a)}{(t-a)P_{r-1}^2(t)}\right]_{t=a} = \frac{P_{r-1}^2(t)}{P_{r-1}^2(a)},$$

$$\begin{aligned} H_{0j}(t) &= \frac{(t-a)P_{r-1}^2(t)}{(t-a_j)^2}\left\{\left[\frac{(t-a_j)^2}{(t-a)P_{r-1}^2(t)}\right]_{t=a_j} + \left[\frac{(t-a_j)^2}{(t-a)P_{r-1}^2(t)}\right]'_{t=a_j}(t-a_j)\right\} \\ &= \frac{(t-a)P_{r-1}^2(t)}{(t-a_j)^2\left[P'_{r-1}(a_j)\right]^2(a_j-a)}\left(1 - \frac{P'_{r-1}(a_j)+(a_j-a)P''_{r-1}(a_j)}{(a_j-a)P'_{r-1}(a_j)}(t-a_j)\right) \end{aligned}$$

and

$$H_{1j}(t) = \frac{(t-a)P_{r-1}^2(t)}{(t-a_j)}\left[\frac{(t-a_j)^2}{(t-a)P_{r-1}^2(t)}\right]_{t=a_j} = \frac{(t-a)P_{r-1}^2(t)}{(t-a_j)(a_j-a)\left[P'_{r-1}(a_j)\right]^2}.$$

Coefficients $\bar{H}_{0j}$, $\bar{H}_{1j}$ and $\bar{H}_{0r}$ we get similarly. □

Theorem 3.3 *Let H_{ij} and $\bar{H}_{ij}$ are defined on $[a,b]$ by (3.12) and (3.13) respectively. Then, if $b_1 = a$, $b_{\bar{r}} = b$, $k_j = 1$ for all $j = 1,\ldots,r$, $l_j = 1$ for all $j = 2,\ldots,\bar{r}-1$ and $l_1 = l_{\bar{r}} = 0$, we have*

$$\begin{aligned} &\ln a\tilde{I}_{\bar{H}_{01}}(\mathbf{p},\mathbf{q}) + \sum_{j=2}^{\bar{r}-1}\left[\ln b_j\tilde{I}_{\bar{H}_{0j}}(\mathbf{p},\mathbf{q}) + \frac{1}{b_j}\tilde{I}_{\bar{H}_{1j}}(\mathbf{p},\mathbf{q})\right] + \ln b\tilde{I}_{\bar{H}_{0r}}(\mathbf{p},\mathbf{q}) \\ &\leq \tilde{I}_{\ln}(\mathbf{p},\mathbf{q}) \leq \sum_{j=1}^{r}\left[\ln a_j\tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + \frac{1}{a_j}\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q})\right] \end{aligned} \tag{3.17}$$

and

$$\begin{aligned} &\sum_{j=1}^{r}\left[a_j\ln a_j\tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + (\ln a_j+1)\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q})\right] \leq D(\mathbf{p}||\mathbf{q}) \\ &\leq a\ln a\tilde{I}_{\bar{H}_{01}}(\mathbf{p},\mathbf{q}) + \sum_{j=2}^{\bar{r}-1}\left[b_j\ln b_j\tilde{I}_{\bar{H}_{0j}}(\mathbf{p},\mathbf{q}) + (\ln b_j+1)\tilde{I}_{\bar{H}_{1j}}(\mathbf{p},\mathbf{q})\right] + b\ln b\tilde{I}_{\bar{H}_{0r}}(\mathbf{p},\mathbf{q}), \end{aligned} \tag{3.18}$$

where

$$H_{0j}(t) = \frac{P_r^2(t)}{(t-a_j)^2\,[P'_r(a_j)]^2}\left(1 - \frac{P''_r(a_j)}{P'_r(a_j)}(t-a_j)\right),$$

$$H_{1j}(t) = \frac{P_r^2(t)}{(t-a_j)\,[P'_r(a_j)]^2},$$

$$\bar{H}_{01}(t) = \frac{(b-t)\bar{P}_{r-1}^2(t)}{(b-a)P_{r-1}^2(a)},$$

$$\bar{H}_{0j}(t) = \frac{(t-a)(b-t)\bar{P}_{r-1}^2(t)}{(b_j-a)(b-b_j)(t-b_j)^2\left[\bar{P}'_{r-1}(b_j)\right]^2}$$
$$\times\left(1+\frac{(2b_j-a-b)\bar{P}'_{r-1}(b_j)-(b-b_j)(b_j-a)\bar{P}''_{r-1}(b_j)}{(b-b_j)(b_j-a)\bar{P}'_{r-1}(b_j)}(t-b_j)\right),$$

$$\bar{H}_{1j}(t) = \frac{(t-a)(b-t)\bar{P}_{r-1}^2(t)}{(t-b_j)(b_j-a)(b-b_j)\left[\bar{P}'_{r-1}(b_j)\right]^2},$$

$$\bar{H}_{0(r+1)}(t) = \frac{(t-a)\bar{P}_{r-1}^2(t)}{(b-a)\bar{P}_{r-1}^2(b)}$$

and

$$P_r(t) = \prod_{j=1}^{r}(t-a_j),\quad \bar{P}_{r-1}(t) = \prod_{j=2}^{r}(t-b_j)$$

for $a \le a_1 < a_2 < \ldots < a_r \le b$, $a < b_2 < \ldots < b_r < b, (r \ge 2)$.

Proof. We use Corollary 3.1 and Corollary 3.4(b) for $n = 2r$ and then calculate

$$\begin{aligned} H_{0j} &= \frac{P_r^2(t)}{(t-a_j)^2}\left\{\left[\frac{(t-a_j)^2}{P_r^2(t)}\right]_{t=a_j}+\left[\frac{(t-a_j)^2}{P_r^2(t)}\right]'_{t=a_j}(t-a_j)\right\} \\ &= \frac{P_r^2(t)}{(t-a_j)^2\left[P'_r(a_j)\right]^2}\left(1-\frac{P''_r(a_j)}{P'_r(a_j)}(t-a_j)\right), \end{aligned}$$

$$H_{1j}(t) = \frac{P_r^2(t)}{t-a_j}\left[\frac{(t-a_j)^2}{P_r^2(t)}\right]_{t=a_j} = \frac{P_r^2(t)}{(t-a_j)\left[P'_r(a_j)\right]^2},$$

$$\bar{H}_{01}(t) = \frac{(t-a)(t-b)\bar{P}_{r-1}^2(t)}{t-a}\left[\frac{t-a}{(t-a)(t-b)\bar{P}_{r-1}^2(t)}\right]_{t=a} = \frac{(b-t)\bar{P}_{r-1}^2(t)}{(b-a)\bar{P}_{r-1}^2(a)},$$

$$\begin{aligned} \bar{H}_{0j}(t) &= \frac{(t-a)(t-b)\bar{P}_{r-1}^2(t)}{(t-b_j)^2} \\ &\times\left\{\left[\frac{(t-b_j)^2}{(t-a)(t-b)P_{r-1}^2(t)}\right]_{t=b_j}+\left[\frac{(t-b_j)^2}{(t-a)(t-b)P_{r-1}^2(t)}\right]'_{t=b_j}(t-b_j)\right\} \\ &= \frac{(t-a)(b-t)\bar{P}_{r-1}^2(t)}{(b_j-a)(b-b_j)(t-b_j)^2\left[\bar{P}'_{r-1}(b_j)\right]^2} \\ &\times\left(1+\frac{(2b_j-a-b)\bar{P}'_{r-1}(b_j)-(b-b_j)(b_j-a)\bar{P}''_{r-1}(b_j)}{(b-b_j)(b_j-a)P'_{r-1}(b_j)}(t-b_j)\right), \end{aligned}$$

$$\begin{aligned}\bar{H}_{1j}(t) &= \frac{(t-a)(t-b)\bar{P}^2_{r-1}(t)}{(t-b_j)}\left[\frac{(t-b_j)^2}{(t-a)(t-b)P^2_{r-1}(t)}\right]_{t=b_j}\\ &= \frac{(t-a)(b-t)\bar{P}^2_{r-1}(t)}{(t-b_j)(b_j-a)(b-b_j)\left[\bar{P}'_{r-1}(b_j)\right]^2}.\end{aligned}$$

Coefficient $\bar{H}_{0(r+1)}$ we get similarly as coefficient $\bar{H}_{01}(t)$. □

Corollary 3.5 *If $n = 2m$ and m is odd, we have*

$$\begin{aligned}&\ln a\tilde{I}_{\bar{H}_{01}}(\mathbf{p},\mathbf{q}) + \sum_{i=1}^{m-1}\frac{(-1)^{i-1}(i-1)!}{a^i}\tilde{I}_{\bar{H}_{i1}}(\mathbf{p},\mathbf{q})\\ &+\ln b\tilde{I}_{\bar{H}_{02}}(\mathbf{p},\mathbf{q}) + \sum_{i=1}^{m-1}\frac{(-1)^{i-1}(i-1)!}{b^i}\tilde{I}_{\bar{H}_{i2}}(\mathbf{p},\mathbf{q})\\ &\le \tilde{I}_{\ln}(\mathbf{p},\mathbf{q}) \le \sum_{j=1}^{m}\left[\ln a_j\tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + \frac{1}{a_j}\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q})\right]\end{aligned} \tag{3.19}$$

and

$$\sum_{j=1}^{m}\left[a_j\ln a_j\tilde{I}_{H_{0j}}(\mathbf{p},\mathbf{q}) + (\ln a_j+1)\tilde{I}_{H_{1j}}(\mathbf{p},\mathbf{q})\right] \le D(\mathbf{p}||\mathbf{q}) \tag{3.20}$$

$$\begin{aligned}&\le a\ln a\tilde{I}_{\bar{H}_{01}}(\mathbf{p},\mathbf{q}) + (\ln a+1)\tilde{I}_{\bar{H}_{11}}(\mathbf{p},\mathbf{q}) + \sum_{i=2}^{m-1}\frac{(-1)^{i-2}(i-2)!}{a^{i-1}}\tilde{I}_{\bar{H}_{i1}}(\mathbf{p},\mathbf{q})\\ &+b\ln b\tilde{I}_{\bar{H}_{02}}(\mathbf{p},\mathbf{q}) + (\ln b+1)\tilde{I}_{\bar{H}_{12}}(\mathbf{p},\mathbf{q}) + \sum_{i=1}^{m-1}\frac{(-1)^{i-2}(i-2)!}{b^{i-2}}\tilde{I}_{\bar{H}_{i2}}(\mathbf{p},\mathbf{q}),\end{aligned}$$

where H_{0j} and H_{1j} as in Theorem 3.3 with $r = m$,

$$\bar{H}_{i1}(t) = \frac{(t-a)^i(t-b)^m}{i!}\sum_{k=0}^{m-1-i}\frac{(-1)^k(m+k-1)!}{[(m-1)!]^2(a-b)^{m+k}}(t-a)^k,$$

and

$$\bar{H}_{i2}(t) = \frac{(t-a)^m(t-b)^i}{i!}\sum_{k=0}^{m-1-i}\frac{(-1)^k(m+k-1)!}{[(m-1)!]^2(b-a)^{m+k}}(t-b)^k.$$

If m even, the inequalities are reversed.

Proof. We use Remark 3.1 and 3.2 and then calculate

$$\begin{aligned}\bar{H}_{i1}(t) &= \frac{1}{i!}\frac{(t-a)^m(t-b)^m}{(t-a)^{m-i}}\sum_{k=0}^{m-1-i}\frac{1}{(m-1)!}\left[\frac{(t-a)^m}{(t-a)^m(t-b)^m}\right]^{(k)}_{t=a}(t-a)^k\\ &= \frac{(t-a)^i(t-b)^m}{i!}\sum_{k=0}^{m-1-i}\frac{(-1)^k(m+k-1)!}{[(m-1)!]^2(a-b)^{m+k}}(t-a)^k.\end{aligned}$$

Coefficient $\bar{H}_{i2}(t)$ we get similarly. □

Remark 3.4 Similarly as in Remark 3.3 we get inequalities of Hermite-Hadamard type related for Shannon entropy.

3.4 Inequalities by using the Zipf-Mandelbrot law

Definition 3.5 *Zipf-Mandelbrot law is a discrete probability distribution depends on three parameters $N \in \{1,2,\ldots\}$, $q \in [0,\infty)$ and $s > 0$, and it is defined by*

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

where

$$H_{N,q,s} := \sum_{k=1}^{N} \frac{1}{(k+q)^s}.$$

If $q = 0$, then Zipf-Mandelbrot law becomes Zipf's law.

Zipf's law is one of the basic laws in information science and bibliometrics. Zipf's law is concerning the frequency of words in the text. We count the number of times each word appears in the text. Words are ranked (r) according to the frequency of occurrence (f). The product of these two numbers is a constant: $r \cdot f = c$.

Apart from the use of this law in bibliometrics and information science, Zipf's law is frequently used in linguistics (see [5], p. 167). In economics and econometrics, this distribution is known as Pareto's law which analyze the distribution of the wealthiest members of the community (see [5], p. 125). These two laws are the same in the mathematical sense, they are only applied in a different context (see [7], p. 294).

The same type of distribution that we have in Zipf's and Pareto's law can be also found in other scientific disciplines, such as: physics, biology, earth and planetary sciences, computer science, demography and the social sciences. For example, the same type of distribution, which we also call the Power law, we can analyze the number of hits on web sites, the magnitude of earthquakes, diameter of moon craters, intensity of solar flares, intensity of wars, population of cities, and others (see [11]).

More general model introduced Benoit Mandelbrot (see [9]), by using arguments on the fractal structure of lexical trees.

The are also quite different interpretation of Zipf-Mandelbrot law in ecology, as it is pointed out in [10] (see also [8] and [12]).

We illustrate our results by using Zipf-Mandelbrot law.

Remark 3.5 Let $\mathbf{q}$ be the Zipf-Mandelbrot law as in Definition 3.5. By applying Theorem 3.2, we have:

$$\ln a \sum_{i=1}^{N} f(i;N,q,s) H_{01}\left(\tfrac{1}{f(i;N,q,s)}\right)$$

$$
\begin{aligned}
&+\sum_{j=2}^{r}\sum_{i=1}^{N} f(i;N,q,s)\left[\ln a_j H_{0j}\left(\tfrac{1}{f(i;N,q,s)}\right)+\frac{1}{a_j}H_{1j}\left(\tfrac{1}{f(i;N,q,s)}\right)\right]\\
\le\ & -\sum_{i=1}^{N} f(i;N,q,s)\ln f(i;N,q,s)\\
\le\ & \sum_{j=1}^{\bar{r}-1}\sum_{i=1}^{N} f(i;N,q,s)\left[\ln b_j \bar{H}_{0j}\left(\tfrac{1}{f(i;N,q,s)}\right)+\frac{1}{b_j}\bar{H}_{1j}\left(\tfrac{1}{f(i;N,q,s)}\right)\right]\\
&+\ln b\sum_{i=1}^{N} f(i;N,q,s)\bar{H}_{0r}\left(\tfrac{1}{f(i;N,q,s)}\right).
\end{aligned}
$$

Let $\mathbf{p_1}$ and $\mathbf{p_2}$ be the Zipf-Mandelbrot law with parameters $N \in \{1,2,\ldots\}$, $q_1, q_2 \in [0,\infty)$ and $s_1, s_2 > 0$, respectively. By applying Theorem 3.2, we have:

$$
\begin{aligned}
&\sum_{j=1}^{\bar{r}-1}\sum_{i=1}^{N} f(i;N,q_2,s_2)\left[b_j\ln b_j \bar{H}_{0j}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)+(\ln b_j+1)\bar{H}_{1j}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)\right]\\
&+b\ln b\sum_{i=1}^{N} f(i;N,q_2,s_2)\bar{H}_{0r}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)\\
\le\ & \sum_{i=1}^{N} f(i;N,q_1,s_1)\ln\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)\\
\le\ & a\ln a\sum_{i=1}^{N} f(i;N,q_2,s_2)H_{01}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)\\
&+\sum_{j=2}^{r}\sum_{i=1}^{N} f(i;N,q_2,s_2)\left[a_j\ln a_j H_{0j}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)+(\ln a_j+1)H_{1j}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)\right].
\end{aligned}
$$

Remark 3.6 Let $\mathbf{q}$ be the Zipf-Mandelbrot law as in Definition 3.5. By applying Theorem 3.3, we have:

$$
\begin{aligned}
&\ln a\sum_{i=1}^{N} f(i;N,q,s)\bar{H}_{01}\left(\tfrac{1}{f(i;N,q,s)}\right)+\sum_{j=2}^{\bar{r}-1}\sum_{i=1}^{N} f(i;N,q,s)\left[\ln b_j \bar{H}_{0j}\left(\tfrac{1}{f(i;N,q,s)}\right)\right.\\
&\left.+\frac{1}{b_j}\bar{H}_{1j}\left(\tfrac{1}{f(i;N,q,s)}\right)\right]+\ln b\sum_{i=1}^{N} f(i;N,q,s)\bar{H}_{0r}\left(\tfrac{1}{f(i;N,q,s)}\right)\\
\le\ & -\sum_{i=1}^{N} f(i;N,q,s)\ln f(i;N,q,s)\\
\le\ & \sum_{j=1}^{r}\sum_{i=1}^{N} f(i;N,q,s)\left[\ln a_j H_{0j}\left(\tfrac{1}{f(i;N,q,s)}\right)+\frac{1}{a_j}H_{1j}\left(\tfrac{1}{f(i;N,q,s)}\right)\right].
\end{aligned}
$$

Let $\mathbf{p_1}$ and $\mathbf{p_2}$ be the Zipf-Mandelbrot law with parameters $N \in \{1,2,\ldots\}$, $q_1, q_2 \in$

$[0,\infty)$ and $s_1, s_2 > 0$, respectively. By applying Theorem 3.3, we have:

$$\sum_{j=1}^{r}\sum_{i=1}^{N} f(i;N,q_2,s_2)\left[a_j \ln a_j H_{0j}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right) + (\ln a_j + 1) H_{1j}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)\right]$$
$$\leq \sum_{i=1}^{N} f(i;N,q_1,s_1) \ln\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)$$
$$\leq a \ln a \sum_{i=1}^{N} f(i;N,q_2,s_2) \bar{H}_{01}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)$$
$$+ \sum_{j=2}^{\bar{r}-1}\sum_{i=1}^{N} f(i;N,q_2,s_2)\left[b_j \ln b_j \bar{H}_{0j}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right) + (\ln b_j + 1)\bar{H}_{1j}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right)\right]$$
$$+ b \ln b \sum_{i=1}^{N} f(i;N,q_2,s_2) \bar{H}_{0r}\left(\tfrac{f(i;N,q_1,s_1)}{f(i;N,q_2,s_2)}\right).$$

Bibliography

[1] R. P. Agarwal, P. J. Y.Wong, *Error Inequalities in Polynomial Interpolation and Their Applications*, Kluwer Academic Publishers, Dordrecht/Boston/London, 1993.

[2] N. S. Barnett, P. Cerone, S. S. Dragomir, A. Sofo, *Approximating Csiszár f-divergence by the use of Taylor's formula with integral remainder*, Math. Ineq. Appl. 5(3) (2002), 417–434.

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

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

[5] V. Diodato, *Dictionary of Bibliometrics*, New York, Haworth Press, 1994.

[6] S. S. Dragomir, *A refinement of Jensen's inequality with applications for f-divergence measures*, Taiwan. J. Math. 14(1), (2010), 153–164.

[7] L. Egghe, R. Rousseau, *Introduction to Informetrics*, Quantitative Methods in Library, Documentation and Information Science. New York, Elsevier Science Publishers, 1990.

[8] S. Frontier, *Diversity and structure in aquatic ecosystems*, Mar Biol Ann Rev 1985; 23: 253–312.

[9] B. Mandelbrot, *An informational theory of the statistical structure of language*, In Jackson, W., editor, Communication Theory, New York. Academic Press, 1953., 486–502.

[10] D. Mouillot, A. Lepretre, *Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity*, Environmental Monitoring and Assessment. Springer. 2000; 63: 279–295.

[11] MEJ. Newman, *Power laws*, Pareto distributions and Zipf.s law. arXiv:condmat/0412004

[12] JB. Wilson, *Methods for fitting dominance/diversity curves*, J Veget Sci 1991; 2: 35–46.

Chapter 4

Zipf-Mandelbrot law and superadditivity of the Jensen functional

Neda Lovričević, Đilda Pečarić and Josip Pečarić

Abstract. Superadditivity property of the discrete Jensen functional is brought in relation to the Csiszár divergence functional. Its monotonicity property and specific bounds are observed consequently in the same context. Some of the well known $f-$ divergences, *e.g.* the Kullback-Leibler divergence, the Hellinger distance, the Bhattacharyya coefficient, the χ^2- divergence and the total variation distance are analyzed in a similar way. All obtained inequalities are eventually interpreted in the environment of the Zipf and the Zipf-Mandelbrot law.

2010 Mathematics Subject Classification. 94A15, 94A17, 26D15, 26A51.
Key words and phrases. Superadditivity, Jensen's functional, Zipf's and Zipf-Mandelbrot's law, Csiszár $f-$divergences, Kullback-Leibler divergence.

4.1 Introduction and preliminaries

Superadditivity property of the discrete Jensen functional

$$J(f,\mathbf{x},\mathbf{p}) := \sum_{i=1}^{n} p_i f(x_i) - P_n f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \tag{4.1}$$

was established in [4] as:

$$J(f,\mathbf{x},\mathbf{p}+\mathbf{q}) \geq J(f,\mathbf{x},\mathbf{p}) + J(f,\mathbf{x},\mathbf{q}), \tag{4.2}$$

where f is a convex function on an interval $I \subset \mathbb{R}$, $\mathbf{x} = (x_1,\ldots,x_n) \in I^n$, $n \geq 2$, $\mathbf{p}$ and $\mathbf{q}$ are nonnegative n-tuples, $P_n = \sum_{i=1}^{n} p_i$, $Q_n = \sum_{i=1}^{n} q_i$, $i = 1,\ldots,n$. Superadditivity was accompanied therein by monotonicity property of the same functional, as its consequent result. Namely, if $\mathbf{p}$ and $\mathbf{q}$ are such that $\mathbf{p} \geq \mathbf{q}$, (*i.e.* $p_i \geq q_i$, $i = 1,\ldots,n$) then

$$J(f,\mathbf{x},\mathbf{p}) \geq J(f,\mathbf{x},\mathbf{q}) \geq 0 \tag{4.3}$$

or

$$\begin{aligned}&\sum_{i=1}^{n} p_i f(x_i) - P_n f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right)\\ &\qquad \geq \sum_{i=1}^{n} q_i f(x_i) - Q_n f\left(\frac{1}{Q_n}\sum_{i=1}^{n} q_i x_i\right) \geq 0.\end{aligned} \tag{4.4}$$

In the monograph [19, p. 717] J. E. Pečarić investigated monotonicity property of the discrete Jensen functional from a different point of view and proved it by using Jensen's inequality and its reverse. Superadditivity property that was thoroughly analyzed in [4] served later as the basis for a more generalized approach *e.g.* in [7], [8], [9] with the results suitably summarized in the monograph [10].

On the other hand, we observe $f-$ divergences which measure the distance between two probability distributions using Csiszár's approach [1, 2]. The Csiszár divergence functional is defined by

$$D_f(\mathbf{r},\mathbf{s}) = \sum_{i=1}^{n} s_i f\left(\frac{r_i}{s_i}\right), \tag{4.5}$$

where $\mathbf{r} = (r_1,\ldots,r_n)$ and $\mathbf{s} = (s_1,\ldots,s_n)$ are positive real $n-$tuples and $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ is a convex function.

Csiszár divergence functional (4.5) may also be defined for nonnegative real $n-$tuples $\mathbf{r}$ and $\mathbf{s}$ with undefined expressions interpreted as

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

or even in a more general setting where $f: I \to \mathbb{R}$, $I \subseteq \mathbb{R}$. Still, in all of the results in the sequel we focus on positive real $n-$tuples $\mathbf{r}$ and $\mathbf{s}$ in definition (4.5).
Furthermore, Csiszár divergence functional (4.5) can be interpreted for special choices of the kernel function f. Thus in the case of positive probability distributions $\mathbf{r}$ and $\mathbf{s}$, that is $r_i, s_i \in \langle 0,1]$, for $i = 1,\ldots,n$ with $\sum_{i=1}^n r_i = \sum_{i=1}^n s_i = 1$ it assumes special forms which we recognize as some well known divergences.

The Kullback-Leibler divergence (see [6], [11], [20]) for positive probability distributions $\mathbf{r} = (r_1,\ldots,r_n)$ and $\mathbf{s} = (s_1,\ldots,s_n)$ is defined by

$$KL(\mathbf{r},\mathbf{s}) := \sum_{i=1}^n r_i \log \frac{r_i}{s_i}. \tag{4.6}$$

In the sequel we analyze results for the logarithm function for different positive bases and distinguish the cases for the bases greater and less than 1.

The Hellinger distance between positive probability distributions $\mathbf{r} = (r_1,\ldots,r_n)$ and $\mathbf{s} = (s_1,\ldots,s_n)$ is defined by

$$h(\mathbf{r},\mathbf{s}) := \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^n \left(\sqrt{r_i} - \sqrt{s_i}\right)^2}. \tag{4.7}$$

The Hellinger distance is a metric and is often used in its squared form, *i.e.* as $h^2(\mathbf{r},\mathbf{s}) := \frac{1}{2}\sum_{i=1}^n \left(\sqrt{r_i} - \sqrt{s_i}\right)^2$.

The Bhattacharyya coefficient is an approximate measure of the amount of overlapping between two positive probability distributions and as such can be used to determine their relative closeness. It is defined as

$$B(\mathbf{r},\mathbf{s}) := \sum_{i=1}^n \sqrt{r_i s_i}. \tag{4.8}$$

Furthermore, the χ^2 (chi-square) divergence is defined as

$$\chi^2(\mathbf{r},\mathbf{s}) := \sum_{i=1}^n \frac{(r_i - s_i)^2}{s_i} \tag{4.9}$$

and the total variation distance or statistical distance is given by

$$V(\mathbf{r},\mathbf{s}) := \sum_{i=1}^n |r_i - s_i|. \tag{4.10}$$

One can find an overview of $f-$ divergences *e.g.* in [3].

$f-$ divergences are observed via the Jensen functional and its listed properties as well as the Zipf-Mandelbrot law and its specified form known as the Zipf law.

Philologist George Kingsley Zipf (1902–1950) studied statistical occurrences in different languages and concluded that, if words of a language were sorted in the order of decreasing frequencies of usage, a word's frequency was inversely proportional to its rank or sequence number in the list [18]. Thus the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word *etc*. It was one of the first academic studies of word frequency and was originally prescribed only for linguistics. It was only later that many other disciplines took credit of it: the Pareto law in economy reveals another aspect of it and the "Zipfian distribution" is present in other fields as well: information science, bibliometrics, social sciences *etc.*
Benoit Mandelbrot (1924–2010) generalized the Zipf law in 1966 [16, 17] and gave its improvement for the count of the low-rank words [13]. It is also used in information sciences for the purpose of indexing [5, 22], in ecological field studies [21] and has its role in art when determining the esthetics criteria in music [15]. The Zipf-Mandelbrot law is a discrete probability distribution and is defined by the following probability mass function:

$$f(i;N,v,w) = \frac{1}{(i+w)^v H_{N,v,w}}, \quad i = 1,\ldots,N, \tag{4.11}$$

where

$$H_{N,v,w} = \sum_{k=1}^{N} \frac{1}{(k+w)^v} \tag{4.12}$$

is a generalization of a harmonic number and $N \in \{1,2,\ldots\}$, $v > 0$ and $w \in [0,\infty)$ are parameters.
For finite N and for $w = 0$ the Zipf-Mandelbrot law is simply called the Zipf law. (In particular, if we observe the infinite N and $w = 0$ we actually have the Zeta distribution.)
According to the expressions above, the probability mass function referring to the Zipf law is

$$f(i;N,v) = \frac{1}{i^v \cdot H_{N,v}}, \quad \text{where} \quad H_{N,v} = \sum_{k=1}^{N} \frac{1}{k^v}, \tag{4.13}$$

that is, out of population of N elements the frequency of elements of rank i is $f(i;N,v)$, where v is the value of the exponent that characterizes the distribution.

The general main inequalities are obtained for Csiszár divergence functional (4.5) via (4.2) and (4.4) in Section 4.2 as well as for the derived special divergences. These yield further working out in the light of the Zipf-Mandelbrot law and the Zipf law, in Section 4.3. Furthermore, results that are presented here generalize for the most part the results previously obtained in [11] and in [14], which is accentuated wherever they occur.

4.2 $f-$ divergences and superadditivity of the Jensen functional

Csiszár divergence functional (4.5) and superadditivity property (4.2) of the discrete Jensen functional (4.1) are integrated in the following theorem.

Theorem 4.1 *Let $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ be a convex function and $\mathbf{r} = (r_1,\ldots,r_n)$ and $\mathbf{s} = (s_1,\ldots,s_n)$ be positive real $n-$tuples such that $R_n = \sum_{i=1}^n r_i$, $S_n = \sum_{i=1}^n s_i$. Suppose $\mathbf{v} = (v_1,\ldots,v_n)$ is a positive real $n-$tuple such that $V_n = \sum_{i=1}^n v_i$. Then*

$$\sum_{i=1}^{n}(s_i+v_i)\,f\left(\frac{r_i}{s_i}\right)-(S_n+V_n)\,f\left(\frac{1}{S_n+V_n}\sum_{i=1}^{n}(s_i+v_i)\,\frac{r_i}{s_i}\right)$$
$$\geq D_f(\mathbf{r},\mathbf{s})-S_nf\left(\frac{R_n}{S_n}\right)+\sum_{i=1}^{n}v_if\left(\frac{r_i}{s_i}\right)-V_nf\left(\frac{1}{V_n}\sum_{i=1}^{n}v_i\frac{r_i}{s_i}\right). \tag{4.14}$$

If f is a concave function, then reverse inequality holds in (4.14).

Proof. Inequality (4.14) follows from inequality (4.2) via definition (4.1) if x_i is replaced by $\frac{r_i}{s_i}$ and p_i replaced by s_i, where $D_f(\mathbf{r},\mathbf{s})$ is the Csiszár functional defined by (4.5). Inequality changes its sign in case of concavity of the function f as a consequence of the Jensen inequality implicitly included. □

Superadditivity of regarding Jensen functional (4.1) yields its monotonicity which reflects on Csiszár functional (4.5) in the following way.

Corollary 4.1 *Let $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ be a convex function and $\mathbf{r} = (r_1,\ldots,r_n)$ and $\mathbf{s} = (s_1,\ldots,s_n)$ be positive real $n-$tuples such that $R_n = \sum_{i=1}^n r_i$, $S_n = \sum_{i=1}^n s_i$. Suppose $\mathbf{t} = (t_1,\ldots,t_n)$ and $\mathbf{u} = (u_1,\ldots,u_n)$ are positive real $n-$tuples such that $T_n = \sum_{i=1}^n t_i$ and $U_n = \sum_{i=1}^n u_i$. If $s_i \geq u_i$, for $i = 1,\ldots,n$ then*

$$D_f(\mathbf{r},\mathbf{s}) \geq S_nf\left(\frac{R_n}{S_n}\right)+\sum_{i=1}^{n}u_if\left(\frac{r_i}{s_i}\right)-U_nf\left(\frac{1}{U_n}\sum_{i=1}^{n}u_i\frac{r_i}{s_i}\right). \tag{4.15}$$

If $s_i \leq t_i$, for $i = 1,\ldots,n$ then

$$D_f(\mathbf{r},\mathbf{s}) \leq S_nf\left(\frac{R_n}{S_n}\right)+\sum_{i=1}^{n}t_if\left(\frac{r_i}{s_i}\right)-T_nf\left(\frac{1}{T_n}\sum_{i=1}^{n}t_i\frac{r_i}{s_i}\right). \tag{4.16}$$

If f is a concave function, then reverse inequalities hold in (4.15) *and* (4.16).

Proof. Inequality (4.15) is obtained from superadditivity (4.2) or (4.14) as we write:

$$\begin{aligned}\sum_{i=1}^{n} s_i f\left(\frac{r_i}{s_i}\right) - S_n f\left(\frac{R_n}{S_n}\right) \geq & \sum_{i=1}^{n} (s_i - u_i) f\left(\frac{r_i}{s_i}\right) \\ & - (S_n - U_n) f\left(\frac{1}{S_n - U_n}\sum_{i=1}^{n}(s_i - u_i)\frac{r_i}{s_i}\right) \\ & + \sum_{i=1}^{n} u_i f\left(\frac{r_i}{s_i}\right) - U_n f\left(\frac{1}{U_n}\sum_{i=1}^{n} u_i \frac{r_i}{s_i}\right), \end{aligned} \tag{4.17}$$

wherefrom we get inequality that corresponds to (4.15):

$$\sum_{i=1}^{n} s_i f\left(\frac{r_i}{s_i}\right) - S_n f\left(\frac{R_n}{S_n}\right) \geq \sum_{i=1}^{n} u_i f\left(\frac{r_i}{s_i}\right) - U_n f\left(\frac{1}{U_n}\sum_{i=1}^{n} u_i \frac{r_i}{s_i}\right), \tag{4.18}$$

since $\sum_{i=1}^{n} (s_i - u_i) f\left(\frac{r_i}{s_i}\right) - (S_n - U_n) f\left(\frac{1}{S_n - U_n}\sum_{i=1}^{n}(s_i - u_i)\frac{r_i}{s_i}\right) \geq 0$ in the observed case of $s_i \geq u_i$, $i = 1, \ldots, n$. On the other side,

$$\begin{aligned}\sum_{i=1}^{n} s_i f\left(\frac{r_i}{s_i}\right) - S_n f\left(\frac{R_n}{S_n}\right) \geq & \sum_{i=1}^{n} (s_i - t_i) f\left(\frac{r_i}{s_i}\right) \\ & - (S_n - T_n) f\left(\frac{1}{S_n - T_n}\sum_{i=1}^{n}(s_i - t_i)\frac{r_i}{s_i}\right) \\ & + \sum_{i=1}^{n} t_i f\left(\frac{r_i}{s_i}\right) - T_n f\left(\frac{1}{T_n}\sum_{i=1}^{n} t_i \frac{r_i}{s_i}\right), \end{aligned} \tag{4.19}$$

wherefrom we get inequality that corresponds to (4.16):

$$\sum_{i=1}^{n} s_i f\left(\frac{r_i}{s_i}\right) - S_n f\left(\frac{R_n}{S_n}\right) \leq \sum_{i=1}^{n} t_i f\left(\frac{r_i}{s_i}\right) - T_n f\left(\frac{1}{T_n}\sum_{i=1}^{n} t_i \frac{r_i}{s_i}\right), \tag{4.20}$$

since $\sum_{i=1}^{n} (s_i - t_i) f\left(\frac{r_i}{s_i}\right) - (S_n - T_n) f\left(\frac{1}{S_n - T_n}\sum_{i=1}^{n}(s_i - t_i)\frac{r_i}{s_i}\right) \leq 0$ in the observed case of $s_i \leq t_i$, $i = 1, \ldots n$. □

Corollary 4.1 was established as the main result in [14] when it was deduced directly from monotonicity property (4.4) similarly as it had been done with the Jensen functional in [19].

Remark 4.1 Inequalities (4.15) and (4.16) are a generalization of specific bounds for Csiszár functional (4.5) that were previously obtained in [11]. Namely, by means of simultaneous inserting the constant $n-$tuples $\mathbf{u}$ and $\mathbf{t}$ into inequalities (4.15) and (4.16), where

$u_i = \min_{i=1,\dots,n}\{s_i\}$ and $t_i = \max_{i=1,\dots,n}\{s_i\}$, we get the following bounds as in [11]:

$$S_n f\left(\frac{R_n}{S_n}\right) + \max_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^n f\left(\frac{r_i}{s_i}\right) - nf\left(\frac{1}{n}\sum_{i=1}^n \frac{r_i}{s_i}\right)\right) \geq D_f(\mathbf{r},\mathbf{s})$$
$$\geq S_n f\left(\frac{R_n}{S_n}\right) + \min_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^n f\left(\frac{r_i}{s_i}\right) - nf\left(\frac{1}{n}\sum_{i=1}^n \frac{r_i}{s_i}\right)\right). \tag{4.21}$$

In the following theorem we establish a similar relation between Csiszár divergence functional (4.5) and superadditivity property (4.2) of discrete Jensen functional (4.1).

Theorem 4.2 *Let $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ be such that $t \mapsto tf(t)$ is a convex function. Assume $\mathbf{r} = (r_1,\dots,r_n)$ and $\mathbf{s} = (s_1,\dots,s_n)$ to be positive real $n-$tuples such that $R_n = \sum_{i=1}^n r_i$, $S_n = \sum_{i=1}^n s_i$. Suppose $\mathbf{v} = (v_1,\dots,v_n)$ is a positive real $n-$tuple such that $V_n = \sum_{i=1}^n v_i$. Then*

$$\sum_{i=1}^n (s_i+v_i)\frac{r_i}{s_i} f\left(\frac{r_i}{s_i}\right) - \left(\sum_{i=1}^n (s_i+v_i)\frac{r_i}{s_i}\right) f\left(\frac{1}{S_n+V_n}\sum_{i=1}^n (s_i+v_i)\frac{r_i}{s_i}\right)$$
$$\geq D_{id\cdot f}(\mathbf{r},\mathbf{s}) - R_n f\left(\frac{R_n}{S_n}\right) + \sum_{i=1}^n v_i\frac{r_i}{s_i} f\left(\frac{r_i}{s_i}\right) - \left(\sum_{i=1}^n v_i\frac{r_i}{s_i}\right) f\left(\frac{1}{V_n}\sum_{i=1}^n v_i\frac{r_i}{s_i}\right). \tag{4.22}$$

where $D_{id\cdot f}(\mathbf{r},\mathbf{s}) := \sum_{i=1}^n r_i f\left(\frac{r_i}{s_i}\right)$.
If $t \mapsto tf(t)$ is a concave function, then reverse inequality holds in (4.22).

Proof. Inequality (4.22) follows from inequality (4.2) via definition (4.1) for convex function $t \mapsto tf(t)$ if x_i is replaced by $\frac{r_i}{s_i}$ and p_i replaced by s_i. Functional $D_{id\cdot f}(\mathbf{r},\mathbf{s})$ is deduced from Csiszár functional (4.5), also for convex function $t \mapsto tf(t)$. Inequality reverses in case of concavity of the function $t \mapsto tf(t)$ as a consequence of the Jensen inequality implicitly included. □

Similarly as the previous one, Corollary 4.2 was also established as the main result in [14] where it was deduced directly from monotonicity property (4.4).

Corollary 4.2 *Let $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ be such that $t \mapsto tf(t)$ is a convex function. Assume $\mathbf{r} = (r_1,\dots,r_n)$ and $\mathbf{s} = (s_1,\dots,s_n)$ to be positive real $n-$tuples such that $R_n = \sum_{i=1}^n r_i$, $S_n = \sum_{i=1}^n s_i$. Suppose $\mathbf{t} = (t_1,\dots,t_n)$ and $\mathbf{u} = (u_1,\dots,u_n)$ are positive real $n-$tuples such that $T_n = \sum_{i=1}^n t_i$ and $U_n = \sum_{i=1}^n u_i$. If $s_i \geq u_i$, for $i = 1,\dots,n$ then*

$$D_{id\cdot f}(\mathbf{r},\mathbf{s}) \geq R_n f\left(\frac{R_n}{S_n}\right) + \sum_{i=1}^n u_i\frac{r_i}{s_i} f\left(\frac{r_i}{s_i}\right) - \left(\sum_{i=1}^n u_i\frac{r_i}{s_i}\right) f\left(\frac{1}{U_n}\sum_{i=1}^n u_i\frac{r_i}{s_i}\right). \tag{4.23}$$

$$D_{id\cdot f}(\mathbf{r},\mathbf{s}) \leq R_n f\left(\frac{R_n}{S_n}\right) + \sum_{i=1}^n t_i\frac{r_i}{s_i} f\left(\frac{r_i}{s_i}\right) - \left(\sum_{i=1}^n t_i\frac{r_i}{s_i}\right) f\left(\frac{1}{T_n}\sum_{i=1}^n t_i\frac{r_i}{s_i}\right). \tag{4.24}$$

If $t \mapsto tf(t)$ is a concave function, then reverse inequalities hold in (4.23) *and* (4.24).

Proof. Follows the same lines as in Corollary 4.1 when we introduce the result from Theorem 4.2 regarding functional $D_{id\cdot f}(\mathbf{r},\mathbf{s})$. $\square$

Remark 4.2 Inequalities (4.23) and (4.24) are a generalization of specific bounds for the functional $D_{id\cdot f}(\mathbf{r},\mathbf{s})$ that were previously obtained in [11]. Namely, by means of simultaneous inserting the constant $n-$tuples $\mathbf{u}$ and $\mathbf{t}$ into inequalities (4.23) and (4.24), where $u_i = \min\limits_{i=1,\dots,n}\{s_i\}$ and $t_i = \max\limits_{i=1,\dots,n}\{s_i\}$, we get the following bounds as in [11]:

$$\begin{aligned} R_n f\left(\frac{R_n}{S_n}\right) + \max_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^{n}\frac{r_i}{s_i}f\left(\frac{r_i}{s_i}\right) - \sum_{i=1}^{n}\frac{r_i}{s_i}f\left(\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}\right)\right) \geq D_{id\cdot f}(\mathbf{r},\mathbf{s}) \\ \geq R_n f\left(\frac{R_n}{S_n}\right) + \min_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^{n}\frac{r_i}{s_i}f\left(\frac{r_i}{s_i}\right) - \sum_{i=1}^{n}\frac{r_i}{s_i}f\left(\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}\right)\right). \end{aligned} \tag{4.25}$$

Corollary 4.2 allows a specific implementation for Kullback-Leibler divergence (4.6).

Corollary 4.3 *Let* $\mathbf{r}$, $\mathbf{s}$ *and* $\mathbf{v}$ *be as in Theorem 4.2. If the logarithm base is greater than 1, then*

$$\begin{aligned} &\sum_{i=1}^{n}(s_i+v_i)\frac{r_i}{s_i}\log\frac{r_i}{s_i} - \left(\sum_{i=1}^{n}(s_i+v_i)\frac{r_i}{s_i}\right)\log\left(\frac{1}{S_n+V_n}\sum_{i=1}^{n}(s_i+v_i)\frac{r_i}{s_i}\right) \\ &\geq \sum_{i=1}^{n} r_i\log\frac{r_i}{s_i} - R_n\log\frac{R_n}{S_n} + \sum_{i=1}^{n} v_i\frac{r_i}{s_i}\log\frac{r_i}{s_i} - \left(\sum_{i=1}^{n} v_i\frac{r_i}{s_i}\right)\log\left(\frac{1}{V_n}\sum_{i=1}^{n} v_i\frac{r_i}{s_i}\right). \end{aligned} \tag{4.26}$$

If the logarithm base is less than 1, then the inequality sign is reversed.

Proof. Follows from Theorem 4.2 for function $t \mapsto t\log t$ which is convex when the logarithm base is greater than 1 and is concave when the logarithm base is less than 1. $\square$

The following corollary leans on (4.26), but the result was also established in [14] directly from (4.4).

Corollary 4.4 *Let* $\mathbf{r}$ *and* $\mathbf{s}$ *be positive probability distributions, i.e.* $r_i, s_i \in \langle 0,1]$, $\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} s_i = 1$ *and* $\mathbf{t}$ *and* $\mathbf{u}$ *are as in Corollary 4.2. If* $s_i \geq u_i$, *for* $i = 1,\dots,n$ *then*

$$KL(\mathbf{r},\mathbf{s}) \geq \sum_{i=1}^{n} u_i\frac{r_i}{s_i}\log\frac{r_i}{s_i} - \left(\sum_{i=1}^{n} u_i\frac{r_i}{s_i}\right)\log\left(\frac{1}{U_n}\sum_{i=1}^{n} u_i\frac{r_i}{s_i}\right), \tag{4.27}$$

where the logarithm base is greater than 1.
If $s_i \leq t_i$, *for* $i = 1,\dots,n$ *then*

$$KL(\mathbf{r},\mathbf{s}) \leq \sum_{i=1}^{n} t_i\frac{r_i}{s_i}\log\frac{r_i}{s_i} - \left(\sum_{i=1}^{n} t_i\frac{r_i}{s_i}\right)\log\left(\frac{1}{T_n}\sum_{i=1}^{n} t_i\frac{r_i}{s_i}\right), \tag{4.28}$$

where the logarithm base is greater than 1.
If the logarithm base is less than 1, then reverse inequalities hold in (4.27) *and* (4.28).

Proof. Taking into account the assumptions on $\mathbf{r}$ and $\mathbf{s}$, functional $\sum_{i=1}^{n} r_i \log \frac{r_i}{s_i}$ from Corollary 4.3 stands now for Kullback-Leibler divergence (4.6). The proof starts with Corollary 4.3 and follows the lines of the proof of Corollary 4.2 (or Corollary 4.1 given here in detail) for function $t \mapsto \log t$ which is convex when the logarithm base is greater than 1 and is concave when the base is less than 1. □

Remark 4.3 Inequalities (4.27) and (4.28) generalize specific bounds for the Kullback-Leibler divergence which were previously obtained in [11]. Namely, by means of simultaneous inserting the constant $n-$tuples $\mathbf{u}$ and $\mathbf{t}$ into inequalities (4.27) and (4.28), where $u_i = \min_{i=1,\dots,n}\{s_i\}$ and $t_i = \max_{i=1,\dots,n}\{s_i\}$, we get the following bounds as presented in [11]:

$$\begin{aligned}\max_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^{n}\frac{r_i}{s_i}\log\frac{r_i}{s_i}-\sum_{i=1}^{n}\frac{r_i}{s_i}\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}\right)\right) &\geq KL(\mathbf{r},\mathbf{s})\\ \geq \min_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^{n}\frac{r_i}{s_i}\log\frac{r_i}{s_i}-\sum_{i=1}^{n}\frac{r_i}{s_i}\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}\right)\right)&,\end{aligned}\tag{4.29}$$

where the logarithm base is greater than 1. If the logarithm base is less than 1, then the inequality signs are reversed.

We proceed with similar results related to other previously introduced divergences: Hellinger distance (4.7), Bhattacharyya coefficient (4.8), chi-square distance (4.9) and total variation distance (4.10).

Corollary 4.5 *Let* $\mathbf{r}$, $\mathbf{s}$ *and* $\mathbf{v}$ *be as in Theorem 4.1. Then*

$$\begin{aligned}&\frac{1}{2}\sum_{i=1}^{n}(s_i+v_i)\left(\sqrt{\frac{r_i}{s_i}}-1\right)^2-\frac{S_n+V_n}{2}\left(\sqrt{\frac{1}{S_n+V_n}\sum_{i=1}^{n}(s_i+v_i)\frac{r_i}{s_i}}-1\right)^2\\ &\geq \frac{1}{2}\sum_{i=1}^{n}(\sqrt{r_i}-\sqrt{s_i})^2-\frac{S_n}{2}\left(\sqrt{\frac{R_n}{S_n}}-1\right)^2+\frac{1}{2}\sum_{i=1}^{n}v_i\left(\sqrt{\frac{r_i}{s_i}}-1\right)^2-\\ &-\frac{V_n}{2}\left(\sqrt{\frac{1}{V_n}\sum_{i=1}^{n}v_i\frac{r_i}{s_i}}-1\right)^2.\end{aligned}\tag{4.30}$$

Proof. Follows from Theorem 4.1 for convex function $t \mapsto \frac{1}{2}\left(\sqrt{t}-1\right)^2$. □

The following corollary concerning the Hellinger distance $h^2(\mathbf{r},\mathbf{s})$ can also be compared with the corresponding result in [14] where it was obtained only by means of monotonicity property (4.4).

Corollary 4.6 *Let* $\mathbf{r}$ *and* $\mathbf{s}$ *be positive probability distributions, i.e.* $r_i, s_i \in \langle 0,1], \sum_{i=1}^{n} r_i = \sum_{i=1}^{n} s_i = 1$ *and* $\mathbf{t}$ *and* $\mathbf{u}$ *are as in Corollary 4.1. If* $s_i \geq u_i$*, for* $i = 1,\dots,n$ *then*

$$h^2(\mathbf{r},\mathbf{s}) \geq \frac{S_n}{2}\left(\sqrt{\frac{R_n}{S_n}}-1\right)^2+\frac{1}{2}\sum_{i=1}^{n}u_i\left(\sqrt{\frac{r_i}{s_i}}-1\right)^2-\frac{U_n}{2}\left(\sqrt{\frac{1}{U_n}\sum_{i=1}^{n}u_i\frac{r_i}{s_i}}-1\right)^2. \tag{4.31}$$

If $s_i \le t_i$, for $i = 1,\ldots,n$ then

$$h^2(\mathbf{r},\mathbf{s}) \le \frac{S_n}{2}\left(\sqrt{\frac{R_n}{S_n}}-1\right)^2 + \frac{1}{2}\sum_{i=1}^{n} t_i\left(\sqrt{\frac{r_i}{s_i}}-1\right)^2 - \frac{T_n}{2}\left(\sqrt{\frac{1}{T_n}\sum_{i=1}^{n} t_i\frac{r_i}{s_i}}-1\right)^2. \tag{4.32}$$

Proof. Since $\mathbf{r}$ and $\mathbf{s}$ are positive probability distributions, we observe the Hellinger distance $h^2(\mathbf{r},\mathbf{s})$ defined by (4.7) as a specific role of $\frac{1}{2}\sum_{i=1}^{n}(\sqrt{r_i}-\sqrt{s_i})^2$ in (4.30). Making use of (4.30) the proof follows the lines as in Corollary 4.1, for convex function $t \mapsto \frac{1}{2}\left(\sqrt{t}-1\right)^2$. □

Remark 4.4 Inequalities (4.31) and (4.32) generalize by means of the constant $n-$ tuples $\mathbf{u}$ and $\mathbf{t}$, $u_i = \min_{i=1,\ldots,n}\{s_i\}$, $t_i = \max_{i=1,\ldots,n}\{s_i\}$ specific bounds for the Hellinger distance which were previously obtained in [11]:

$$\begin{aligned}
&\frac{1}{2}\max_{i=1,\ldots,n}\{s_i\}\left(\sum_{i=1}^{n}\left(\sqrt{\frac{r_i}{s_i}}-1\right)^2 - n\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}}-1\right)^2\right)\\
&\ge h^2(\mathbf{r},\mathbf{s})\\
&\ge \frac{1}{2}\min_{i=1,\ldots,n}\{s_i\}\left(\sum_{i=1}^{n}\left(\sqrt{\frac{r_i}{s_i}}-1\right)^2 - n\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}}-1\right)^2\right).
\end{aligned} \tag{4.33}$$

Corollary 4.7 *Let* $\mathbf{r}$, $\mathbf{s}$ *and* $\mathbf{v}$ *be as in Theorem 4.1. Then*

$$\begin{aligned}
&(S_n+V_n)\sqrt{\frac{1}{S_n+V_n}\sum_{i=1}^{n}(s_i+v_i)\frac{r_i}{s_i}} - \sum_{i=1}^{n}(s_i+v_i)\sqrt{\frac{r_i}{s_i}}\\
&\ge S_n\sqrt{\frac{R_n}{S_n}} - \sum_{i=1}^{n}\sqrt{r_i s_i} + V_n\sqrt{\frac{1}{V_n}\sum_{i=1}^{n} v_i\frac{r_i}{s_i}} - \sum_{i=1}^{n} v_i\sqrt{\frac{r_i}{s_i}}.
\end{aligned} \tag{4.34}$$

Proof. Follows from Theorem 4.1 for convex function $t \mapsto -\sqrt{t}$. □

Monotonicity property for Bhattacharyya coefficient (4.8) is deduced from superadditivity property (4.34) as follows, although it was formerly deduced in [14] with the help of (4.4).

Corollary 4.8 *Let* $\mathbf{r}$ *and* $\mathbf{s}$ *be positive probability distributions, i.e.* $r_i, s_i \in \langle 0,1]$, $\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} s_i = 1$ *and* $\mathbf{t}$ *and* $\mathbf{u}$ *are as in Corollary 4.1. If* $s_i \ge u_i$, *for* $i = 1,\ldots,n$ *then*

$$B(\mathbf{r},\mathbf{s}) \ge -S_n\sqrt{\frac{R_n}{S_n}} - \sum_{i=1}^{n} u_i\sqrt{\frac{r_i}{s_i}} + U_n\sqrt{\frac{1}{U_n}\sum_{i=1}^{n} u_i\frac{r_i}{s_i}}. \tag{4.35}$$

If $s_i \le t_i$, for $i = 1,\ldots,n$ then

$$B(\mathbf{r},\mathbf{s}) \le -S_n\sqrt{\frac{R_n}{S_n}} - \sum_{i=1}^{n} t_i\sqrt{\frac{r_i}{s_i}} + T_n\sqrt{\frac{1}{T_n}\sum_{i=1}^{n} t_i\frac{r_i}{s_i}}. \tag{4.36}$$

Proof. With positive probability distributions $\mathbf{r}$ and $\mathbf{s}$ involved, we actually deal with the Bhattacharyya coefficient $B(\mathbf{r},\mathbf{s})$ defined by (4.8), when observing functional $-\sum_{i=1}^{n}\sqrt{r_i s_i}$. The proof is carried out by (4.34) analogously as in Corollary 4.1, for convex function $t\mapsto -\sqrt{t}$. □

Remark 4.5 If we make use of constant $n-$tuples $\mathbf{u}$ and $\mathbf{t}$ with $u_i = \min_{i=1,\dots,n}\{s_i\}$ and $t_i = \max_{i=1,\dots,n}\{s_i\}$ and insert them into inequalities (4.35) and (4.36), we get :

$$1-\min_{i=1,\dots,n}\{s_i\}\left(n\sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}}-\sum_{i=1}^{n}\sqrt{\frac{r_i}{s_i}}\right)\geq B(\mathbf{r},\mathbf{s})$$
$$\geq 1-\max_{i=1,\dots,n}\{s_i\}\left(n\sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}}-\sum_{i=1}^{n}\sqrt{\frac{r_i}{s_i}}\right), \tag{4.37}$$

that is, bounds from [11] which are a special case in this more general setting.

Corollary 4.9 *Let* $\mathbf{r}$, $\mathbf{s}$ *and* $\mathbf{v}$ *be as in Theorem 4.1. Then*

$$\sum_{i=1}^{n}(s_i+v_i)\left(\frac{r_i}{s_i}-1\right)^2-(S_n+V_n)\left(\frac{1}{S_n+V_n}\sum_{i=1}^{n}(s_i+v_i)\frac{r_i}{s_i}-1\right)^2$$
$$\geq\sum_{i=1}^{n}\frac{(r_i-s_i)^2}{s_i}-S_n\left(\frac{R_n}{S_n}-1\right)^2+\sum_{i=1}^{n}v_i\left(\frac{r_i}{s_i}-1\right)^2-V_n\left(\frac{1}{V_n}\sum_{i=1}^{n}v_i\frac{r_i}{s_i}-1\right)^2. \tag{4.38}$$

Proof. Follows from Theorem 4.1 for convex function $t\mapsto(t-1)^2$. □

The following corollary on chi-square divergence (4.9) leans on (4.38), but the result was also established in [14] from (4.4).

Corollary 4.10 *Let* $\mathbf{r}$ *and* $\mathbf{s}$ *be positive probability distributions, i.e.* $r_i, s_i\in\langle 0,1]$, $\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} s_i = 1$ *and* $\mathbf{t}$ *and* $\mathbf{u}$ *are as in Corollary 4.1. If* $s_i\geq u_i$, *for* $i=1,\dots,n$ *then*

$$\chi^2(\mathbf{r},\mathbf{s})\geq\sum_{i=1}^{n}u_i\left(\frac{r_i}{s_i}-1\right)^2-U_n\left(\frac{1}{U_n}\sum_{i=1}^{n}u_i\frac{r_i}{s_i}-1\right)^2. \tag{4.39}$$

If $s_i\leq t_i$, *for* $i=1,\dots,n$ *then*

$$\chi^2(\mathbf{r},\mathbf{s})\leq\sum_{i=1}^{n}t_i\left(\frac{r_i}{s_i}-1\right)^2-T_n\left(\frac{1}{T_n}\sum_{i=1}^{n}t_i\frac{r_i}{s_i}-1\right)^2. \tag{4.40}$$

Proof. When we observe functional $\sum_{i=1}^{n}\frac{(r_i-s_i)^2}{s_i}$ in (4.38) under the additional assumptions on $\mathbf{r}$ and $\mathbf{s}$, then inequalities (4.39) and (4.40) concern the chi-square divergence $\chi^2(\mathbf{r},\mathbf{s})$ defined by (4.9). The proof is carried out by (4.38) analogously as in Corollary 4.1, for convex function $t\mapsto(t-1)^2$. □

Remark 4.6 Inequalities (4.39) and (4.40) generalize specific bounds for the chi-square divergence which had been previously obtained in [11]. Namely, by means of simultaneous inserting the constant $n-$tuples $\mathbf{u}$ and $\mathbf{t}$ into inequalities (4.39) and (4.40), where $u_i = \min_{i=1,\dots,n}\{s_i\}$ and $t_i = \max_{i=1,\dots,n}\{s_i\}$, we get the following bounds as presented in [11]:

$$\begin{aligned}
&\max_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^{n}\left(\frac{r_i}{s_i}-1\right)^2 - n\left(\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}-1\right)^2\right)\\
&\geq \chi^2(\mathbf{r},\mathbf{s})\\
&\geq \min_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^{n}\left(\frac{r_i}{s_i}-1\right)^2 - n\left(\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}-1\right)^2\right).
\end{aligned} \tag{4.41}$$

Corollary 4.11 *Let* $\mathbf{r}$, $\mathbf{s}$ *and* $\mathbf{v}$ *be as in Theorem 4.1. Then*

$$\begin{aligned}
&\sum_{i=1}^{n}(s_i+v_i)\left|\frac{r_i}{s_i}-1\right| - (S_n+V_n)\left|\frac{1}{S_n+V_n}\sum_{i=1}^{n}(s_i+v_i)\frac{r_i}{s_i}-1\right|\\
&\geq \sum_{i=1}^{n}|r_i-s_i| - S_n\left|\frac{R_n}{S_n}-1\right| + \sum_{i=1}^{n}v_i\left|\frac{r_i}{s_i}-1\right| - V_n\left|\frac{1}{V_n}\sum_{i=1}^{n}v_i\frac{r_i}{s_i}-1\right|
\end{aligned} \tag{4.42}$$

Proof. Follows from Theorem 4.1 for convex function $t \mapsto |t-1|$. □

The following corollary concerns the total variation distance $V(\mathbf{r},\mathbf{s})$ and can also be compared with the corresponding result in [14] where it was obtained only by means of monotonicity property (4.4).

Corollary 4.12 *Let* $\mathbf{r}$ *and* $\mathbf{s}$ *be positive probability distributions, i.e.* $r_i, s_i \in \langle 0,1]$, $\sum_{i=1}^{n} r_i = \sum_{i=1}^{n} s_i = 1$ *and* $\mathbf{t}$ *and* $\mathbf{u}$ *are as in Corollary 4.1. If* $s_i \geq u_i$, *for* $i = 1,\dots,n$ *then*

$$V(\mathbf{r},\mathbf{s}) \geq S_n\left|\frac{R_n}{S_n}-1\right| + \sum_{i=1}^{n}u_i\left|\frac{r_i}{s_i}-1\right| - U_n\left|\frac{1}{U_n}\sum_{i=1}^{n}u_i\frac{r_i}{s_i}-1\right|. \tag{4.43}$$

If $s_i \leq t_i$, *for* $i = 1,\dots,n$ *then*

$$V(\mathbf{r},\mathbf{s}) \leq S_n\left|\frac{R_n}{S_n}-1\right| + \sum_{i=1}^{n}t_i\left|\frac{r_i}{s_i}-1\right| - T_n\left|\frac{1}{T_n}\sum_{i=1}^{n}t_i\frac{r_i}{s_i}-1\right|. \tag{4.44}$$

Proof. When $\mathbf{r}$ and $\mathbf{s}$ are observed as positive probability distributions, then functional $\sum_{i=1}^{n}|r_i-s_i|$ is the total variation distance $V(\mathbf{r},\mathbf{s})$ defined by (4.10). The proof follows the lines of the proof of Corollary 4.1, after relation (4.42), for convex function $t \mapsto |t-1|$. □

Remark 4.7 Inequalities (4.43) and (4.44) generalize by means of the constant $n-$ tuples $\mathbf{u}$ and $\mathbf{t}$, $u_i = \min_{i=1,\dots,n}\{s_i\}$, $t_i = \max_{i=1,\dots,n}\{s_i\}$ specific bounds for the total variation distance which were previously obtained in [11]:

$$
\begin{aligned}
\max_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^{n}\left|\frac{r_i}{s_i}-1\right|-n\left|\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}-1\right|\right) &\geq V(\mathbf{r},\mathbf{s}) \\
&\geq \min_{i=1,\dots,n}\{s_i\}\left(\sum_{i=1}^{n}\left|\frac{r_i}{s_i}-1\right|-n\left|\frac{1}{n}\sum_{i=1}^{n}\frac{r_i}{s_i}-1\right|\right). \qquad (4.45)
\end{aligned}
$$

4.3 Zipf-Mandelbrot law in $f-$ divergences

If we define s_i as (4.11), that is as the Zipf-Mandelbrot law probability mass functions $f(i;N,v,w)$, for $i = 1,\dots,N$, we can use a new environment to observe the previously obtained results. When observed with the Zipf-Mandelbrot $N-$tuple $\mathbf{s}$ included, the Csiszár functional $D_f(\mathbf{r},\mathbf{s})$ defined by (4.5) becomes

$$
D_f(i,N,v_2,w_2,\mathbf{r}) = \sum_{i=1}^{N}\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}f\left(r_i\,(i+w_2)^{v_2}H_{N,v_2,w_2}\right), \qquad (4.46)
$$

where $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ and $N \in \mathbb{N}$, $v_2,w_2 > 0$ are parameters.
Csiszár functional (4.5) assumes the following form when $\mathbf{r}$ and $\mathbf{s}$ are both defined as Zipf-Mandelbrot law $N-$tuples:

$$
D_f(i,N,v_1,v_2,w_1,w_2) = \sum_{i=1}^{N}\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right), \qquad (4.47)
$$

where $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ and $N \in \mathbb{N}$, $v_1,v_2,w_1,w_2 > 0$ are parameters.

Finally, both $N-$tuples $\mathbf{r}$ and $\mathbf{s}$ may be defined via the Zipf law (4.13) where $w_1{=}w_2{=}0$ and thus Csiszár functional (4.5) assumes the form:

$$
D_f(i,N,v_1,v_2) = \sum_{i=1}^{N}\frac{1}{i^{v_2}H_{N,v_2}}f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right). \qquad (4.48)
$$

Remark 4.8 For all three types of the Csiszár functional defined via the Zipf and the Zipf-Mandelbrot laws, superadditivity (4.2) provides new types of results related to Theorem 4.1 and Theorem 4.2. All results on monotonicity presented in sequel were deduced previously in [14] from (4.4) and are now deduced from superadditivity, which is a more general aspect.

In the first case, that is for the Csiszár functional $D_f(i,N,v_2,w_2,\mathbf{r})$ given as in (4.46) we transform Theorem 4.1 and Theorem 4.2 in the following way.

Corollary 4.13 *Let $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ be a convex function, v_2, $w_2 > 0$ and $\mathbf{r} = (r_1,\dots,r_N)$, $\overline{\mathbf{v}} = (\overline{v}_1,\dots,\overline{v}_N)$ be positive real $N-$tuples such that $R_N = \sum_{i=1}^N r_i$, $\overline{V}_N = \sum_{i=1}^N \overline{v}_i$. Then*

$$
\begin{aligned}
&\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}+\overline{v}_i\right) f\left(r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right)\\
&-\left(1+\overline{V}_N\right) f\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}+\overline{v}_i\right) r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right)\\
&\geq D_f(i,N,v_2,w_2,\mathbf{r}) - f(R_N) + \sum_{i=1}^{N}\overline{v}_i f\left(r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right)\\
&-\overline{V}_N f\left(\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right).
\end{aligned}
\tag{4.49}
$$

If f is a concave function, then reverse inequality holds in (4.49).
Suppose $t \mapsto tf(t)$ is a convex function. Then

$$
\begin{aligned}
&\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}+\overline{v}_i\right) r_i(i+w_2)^{v_2}H_{N,v_2,w_2} f\left(r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right)\\
&-\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}+\overline{v}_i\right) r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\,\cdot\\
&\cdot\; f\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}+\overline{v}_i\right) r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right)\\
&\geq D_{id\cdot f}(i,N,v_2,w_2,\mathbf{r}) - R_N f(R_N) + \sum_{i=1}^{N}\overline{v}_i r_i(i+w_2)^{v_2}H_{N,v_2,w_2} f\left(r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right)\\
&-\left(\sum_{i=1}^{N}\overline{v}_i r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right) f\left(\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i r_i(i+w_2)^{v_2}H_{N,v_2,w_2}\right),
\end{aligned}
\tag{4.50}
$$

where $D_{id\cdot f}(i,N,v_2,w_2,\mathbf{r}) := \sum_{i=1}^{N} r_i f\left((i+w_2)^{v_2}H_{N,v_2,w_2}\right)$.
If $t \mapsto tf(t)$ is a concave function, then reverse inequality holds in (4.50).

Proof. Inequality (4.49) leans on the proof of Theorem 4.1 wherein we insert for s_i the expression $\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w2}}$ by definition (4.11) of the Zipf-Mandelbrot law and $S_N = 1$. Inequality (4.50) follows analogously after the proof of Theorem 4.2. Inequalities change their signs in the case of concavity of functions f or $t \mapsto tf(t)$ as a consequence of the Jensen inequality implicitly included. □

Monotonicity property of Jensen functional (4.1) reflects on $D_f(i,N,v_2,w_2,\mathbf{r})$ and $D_{id\cdot f}(i,N,v_2,w_2,\mathbf{r})$ as follows.

Corollary 4.14 *Let f, v_2, w_2 and $\mathbf{r}$ be as in Corollary 4.13. Suppose $\mathbf{t} = (t_1, \ldots, t_N)$ and $\mathbf{u} = (u_1, \ldots, u_N)$ are positive real $N-$tuples such that $T_N = \sum_{i=1}^N t_i$ and $U_N = \sum_{i=1}^N u_i$. If $\frac{1}{(i+w_2)^{v_2} H_{N,v_2,w_2}} \geq u_i$, for $i = 1, \ldots, N$ then*

$$\begin{aligned} D_f(i,N,v_2,w_2,\mathbf{r}) \geq\; & f(R_N) + \sum_{i=1}^N u_i f\left(r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right) \\ & - U_N f\left(\frac{1}{U_N}\sum_{i=1}^N u_i r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right). \end{aligned} \tag{4.51}$$

If $\frac{1}{(i+w_2)^{v_2} H_{N,v_2,w_2}} \leq t_i$, for $i = 1, \ldots, N$ then

$$\begin{aligned} D_f(i,N,v_2,w_2,\mathbf{r}) \leq\; & f(R_N) + \sum_{i=1}^N t_i f\left(r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right) \\ & - T_N f\left(\frac{1}{T_N}\sum_{i=1}^N t_i r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right). \end{aligned} \tag{4.52}$$

If f is a concave function, then reverse inequalities hold in (4.51) *and* (4.52).

Suppose $t \mapsto t f(t)$ is a convex function.
If $\frac{1}{(i+w_2)^{v_2} H_{N,v_2,w_2}} \geq u_i$, for $i = 1, \ldots, N$ then

$$\begin{aligned} D_{id\cdot f}(i,N,v_2,w_2,\mathbf{r}) \geq & R_N f(R_N) + \sum_{i=1}^N u_i r_i (i+w_2)^{v_2} H_{N,v_2,w_2} f\left(r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right) \\ & - \left(\sum_{i=1}^N u_i r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right) f\left(\frac{1}{U_N}\sum_{i=1}^N u_i r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right), \end{aligned} \tag{4.53}$$

If $\frac{1}{(i+w_2)^{v_2} H_{N,v_2,w_2}} \leq t_i$, for $i = 1, \ldots, N$ then

$$\begin{aligned} D_{id\cdot f}(i,N,v_2,w_2,\mathbf{r}) \leq & R_N f(R_N) + \sum_{i=1}^N t_i r_i (i+w_2)^{v_2} H_{N,v_2,w_2} f\left(r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right) \\ & - \left(\sum_{i=1}^N t_i r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right) f\left(\frac{1}{T_N}\sum_{i=1}^N t_i r_i (i+w_2)^{v_2} H_{N,v_2,w_2}\right). \end{aligned} \tag{4.54}$$

If $t \mapsto t f(t)$ is a concave function, then reverse inequalities hold in (4.53) *and* (4.54).

Proof. Follows from Corollary 4.13 or is carried out as for Corollary 4.1 and Corollary 4.2. □

Remark 4.9 If we put $u_i = \min_{i=1,\dots,N}\left\{\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}\right\} = \frac{1}{(N+w_2)^{v_2}H_{N,v_2,w_2}}$ and $t_i = \max_{i=1,\dots,N}\left\{\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}\right\} = \frac{1}{(1+w_2)^{v_2}H_{N,v_2,w_2}}$ simultaneously into inequalities (4.51) and (4.52) we get the following bounds as a special case of Corollary 4.14. These were obtained earlier in [11]:

$$\begin{aligned} f(R_N) + \frac{1}{(1+w_2)^{v_2}H_{N,v_2,w_2}}\Lambda_1 &\geq D_f(i,N,v_2,w_2,\mathbf{r}) \\ &\geq f(R_N) + \frac{1}{(N+w_2)^{v_2}H_{N,v_2,w_2}}\,\Lambda_1, \end{aligned} \tag{4.55}$$

where $\Lambda_1 = \sum_{i=1}^{N} f\left(r_i\,(i+w_2)^{v_2}\,H_{N,v_2,w_2}\right) - Nf\left(\frac{1}{N}\sum_{i=1}^{N} r_i\,(i+w_2)^{v_2}\,H_{N,v_2,w_2}\right)$.
If we repeat the similar procedure with inequalities (4.53) and (4.54), we get the analogous bounds for $D_{id\cdot f}(i,N,v_2,w_2,\mathbf{r})$, previously obtained in [11], as well:

$$\begin{aligned} R_N f(R_N) + \frac{1}{(1+w_2)^{v_2}H_{N,v_2,w_2}}\,\widetilde{\Lambda}_1 &\geq D_{id\cdot f}(i,N,v_2,w_2,\mathbf{r}) \\ &\geq R_N f(R_N) + \frac{1}{(N+w_2)^{v_2}H_{N,v_2,w_2}}\,\widetilde{\Lambda}_1, \end{aligned} \tag{4.56}$$

where

$$\begin{aligned} \widetilde{\Lambda}_1 &= \sum_{i=1}^{N} r_i\,(i+w_2)^{v_2}\,H_{N,v_2,w_2} f\left(r_i\,(i+w_2)^{v_2}\,H_{N,v_2,w_2}\right) \\ &\quad - \left(\sum_{i=1}^{N} r_i\,(i+w_2)^{v_2}\,H_{N,v_2,w_2}\right) f\left(\frac{1}{N}\sum_{i=1}^{N} r_i\,(i+w_2)^{v_2}\,H_{N,v_2,w_2}\right). \end{aligned} \tag{4.57}$$

In the second case, that is for the Csiszár functional $D_f(i,N,v_1,v_2,w_1,w_2)$ as in (4.47), we transform Theorem 4.1 and Theorem 4.2 as follows.

Corollary 4.15 *Let $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ be a convex function and $v_1,v_2,\ w_1,w_2 > 0$. Let $\overline{\mathbf{v}} = (\overline{v}_1,\dots,\overline{v}_N)$ be a positive real $N-$tuple such that $\overline{V}_N = \sum_{i=1}^{N}\overline{v}_i$. Then*

$$\begin{aligned} &\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}} + \overline{v}_i\right) f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right) \\ &- \left(1+\overline{V}_N\right) f\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}} + \overline{v}_i\right)\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right) \\ &\geq D_f(i,N,v_1,v_2,w_1,w_2) - f(1) + \sum_{i=1}^{N}\overline{v}_i f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right) \\ &- \overline{V}_N f\left(\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right). \end{aligned} \tag{4.58}$$

If f is a concave function, then reverse inequality holds in (4.58).
Suppose $t \mapsto tf(t)$ is a convex function. Then

$$
\begin{aligned}
&\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}+\overline{v}_i\right)\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)\\
&-\left(\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}+\overline{v}_i\right)\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)\\
&\cdot\ f\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}+\overline{v}_i\right)\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)\\
&\geq D_{id\cdot f}(i,N,v_1,v_2,w_1,w_2)-f(1)\\
&+\ \sum_{i=1}^{N}\overline{v}_i\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)\\
&-\left(\sum_{i=1}^{N}\overline{v}_i\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)f\left(\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right),
\end{aligned}
\tag{4.59}
$$

where $D_{id\cdot f}(i,N,v_1,v_2,w_1,w_2):=\sum_{i=1}^{N}\frac{1}{(i+w_1)^{v_1}H_{N,v_1,w_1}}f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)$.
If $t \mapsto tf(t)$ is a concave function, then reverse inequality holds in (4.59).

Proof. Inequality (4.58) leans on the proof of Theorem 4.1 wherein we insert for r_i and s_i expressions $\frac{1}{(i+w_1)^{v_1}H_{N,v_1,w_1}}$ and $\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}$ by definition (4.11) of the Zipf-Mandelbrot law and $R_N = S_N = 1$. Inequality (4.59) follows analogously after the proof of Theorem 4.2. Inequalities change their signs in the case of concavity of functions f or $t \mapsto tf(t)$ as a consequence of the Jensen inequality implicitly included. $\square$

Monotonicity property of Jensen functional (4.1) reflects on $D_f(i,N,v_1,v_2,w_1,w_2)$ and $D_{id\cdot f}(i,N,v_1,v_2,w_1,w_2)$ as follows.

Corollary 4.16 *Let f, v_1, v_2, w_1 and w_2 be as in Corollary 4.15. Suppose $\mathbf{t}=(t_1,\ldots,t_N)$ and $\mathbf{u}=(u_1,\ldots,u_N)$ are positive real $N-$tuples such that $T_N=\sum_{i=1}^{N}t_i$ and $U_N=\sum_{i=1}^{N}u_i$. If $\frac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}\geq u_i$, for $i=1,\ldots,N$ then*

$$
\begin{aligned}
D_f(i,N,v_1,v_2,w_1,w_2)\geq f(1)+\sum_{i=1}^{N}u_i f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)&\\
-U_N f\left(\frac{1}{U_N}\sum_{i=1}^{N}u_i\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)&.
\end{aligned}
\tag{4.60}
$$

If $\dfrac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}} \le t_i$*, for* $i=1,\dots,N$ *then*

$$D_f(i,N,v_1,v_2,w_1,w_2) \le f(1)+\sum_{i=1}^{N} t_i f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right) -T_N f\left(\frac{1}{T_N}\sum_{i=1}^{N} t_i \frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right). \tag{4.61}$$

If f *is a concave function, then reverse inequalities hold in* (4.60) *and* (4.61).
Suppose $t \mapsto t f(t)$ *is a convex function. If* $\dfrac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}} \ge u_i$*, for* $i=1,\dots,N$ *then*

$$D_{id\cdot f}(i,N,v_1,v_2,w_1,w_2) \ge f(1)+\sum_{i=1}^{N} u_i \frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}} f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right) - \left(\sum_{i=1}^{N} u_i \frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right) f\left(\frac{1}{U_N}\sum_{i=1}^{N} u_i \frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right), \tag{4.62}$$

where $D_{id\cdot f}(i,N,v_1,v_2,w_1,w_2) := \sum_{i=1}^{N} \dfrac{1}{(i+w_1)^{v_1}H_{N,v_1,w_1}} f\left(\dfrac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right)$.
If $\dfrac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}} \le t_i$*, for* $i=1,\dots,N$ *then*

$$D_{id\cdot f}(i,N,v_1,v_2,w_1,w_2) \le f(1)+\sum_{i=1}^{N} t_i \frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}} f\left(\frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right) - \left(\sum_{i=1}^{N} t_i \frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right) f\left(\frac{1}{T_N}\sum_{i=1}^{N} t_i \frac{(i+w_2)^{v_2}H_{N,v_2,w_2}}{(i+w_1)^{v_1}H_{N,v_1,w_1}}\right). \tag{4.63}$$

If $t \mapsto t f(t)$ *is a concave function, then reverse inequalities hold in* (4.62) *and* (4.63).

Proof. Follows from Corollary 4.15 or is carried out as for Corollary 4.1 and Corollary 4.2. □

Remark 4.10 For $u_i = \min_{i=1,\dots,N}\left\{\dfrac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}\right\} = \dfrac{1}{(N+w_2)^{v_2}H_{N,v_2,w_2}}$ and $t_i = \max_{i=1,\dots,N}\left\{\dfrac{1}{(i+w_2)^{v_2}H_{N,v_2,w_2}}\right\} = \dfrac{1}{(1+w_2)^{v_2}H_{N,v_2,w_2}}$ inequalities (4.60) and (4.61) assume the form of the bounds that were obtained earlier in [11] :

$$f(1)+\frac{1}{(1+w_2)^{v_2}H_{N,v_2,w_2}}\Lambda_2 \ge D_f(i,N,v_1,v_2,w_1,w_2) \ge f(1)+\frac{1}{(N+w_2)^{v_2}H_{N,v_2,w_2}}\Lambda_2, \tag{4.64}$$

where $\Lambda_2 = \sum_{i=1}^{N} f\left(\frac{(i+w_2)^{v_2} H_{N,v_2,w_2}}{(i+w_1)^{v_1} H_{N,v_1,w_1}}\right) - Nf\left(\frac{1}{N}\sum_{i=1}^{N} \frac{(i+w_2)^{v_2} H_{N,v_2,w_2}}{(i+w_1)^{v_1} H_{N,v_1,w_1}}\right)$.
If we repeat the similar procedure with inequalities (4.62) and (4.63), we get the analogous bounds for $D_{id\cdot f}(i,N,v_1,v_2,w_1,w_2)$, as in [11]:

$$\begin{aligned} f(1) + \frac{1}{(1+w_2)^{v_2} H_{N,v_2,w_2}} \widetilde{\Lambda}_2 &\geq D_{id\cdot f}(i,N,v_1,v_2,w_1,w_2) \\ &\geq f(1) + \frac{1}{(N+w_2)^{v_2} H_{N,v_2,w_2}} \widetilde{\Lambda}_2, \end{aligned} \tag{4.65}$$

where

$$\begin{aligned} \widetilde{\Lambda}_2 = & \sum_{i=1}^{N} \frac{(i+w_2)^{v_2} H_{N,v_2,w_2}}{(i+w_1)^{v_1} H_{N,v_1,w_1}} f\left(\frac{(i+w_2)^{v_2} H_{N,v_2,w_2}}{(i+w_1)^{v_1} H_{N,v_1,w_1}}\right) \\ & - \left(\sum_{i=1}^{N} \frac{(i+w_2)^{v_2} H_{N,v_2,w_2}}{(i+w_1)^{v_1} H_{N,v_1,w_1}}\right) f\left(\frac{1}{N}\sum_{i=1}^{N} \frac{(i+w_2)^{v_2} H_{N,v_2,w_2}}{(i+w_1)^{v_1} H_{N,v_1,w_1}}\right). \end{aligned} \tag{4.66}$$

Finally, when the Csiszár functional $D_f(i,N,v_1,v_2)$ is defined as in (4.48), that is by means of the Zipf law $N-$tuples, Theorem 4.1 and Theorem 4.2 assume the following forms.

Corollary 4.17 *Let $f\colon \langle 0,\infty\rangle \to \mathbb{R}$ be a convex function and $v_1, v_2 > 0$. Suppose $\overline{\mathbf{v}} = (\overline{v}_1,\ldots,\overline{v}_N)$ be a positive real $N-$tuple such that $\overline{V}_N = \sum_{i=1}^{N} \overline{v}_i$. Then*

$$\begin{aligned} & \sum_{i=1}^{N}\left(\frac{1}{i^{v_2} H_{N,v_2}} + \overline{v}_i\right) f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) \\ & - \left(1+\overline{V}_N\right) f\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{i^{v_2} H_{N,v_2}} + \overline{v}_i\right) i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) \\ & \geq D_f(i,N,v_1,v_2) - f(1) + \sum_{i=1}^{N} \overline{v}_i f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) \\ & - \overline{V}_N f\left(\frac{1}{\overline{V}_N}\sum_{i=1}^{N} \overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right). \end{aligned} \tag{4.67}$$

If f is a concave function, then reverse inequality holds in (4.67).
Suppose $t \mapsto tf(t)$ is a convex function. Then

$$\begin{aligned} & \sum_{i=1}^{N}\left(\frac{1}{i^{v_2} H_{N,v_2}} + \overline{v}_i\right) i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) \\ & - \left(\sum_{i=1}^{N}\left(\frac{1}{i^{v_2} H_{N,v_2}} + \overline{v}_i\right) i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) \end{aligned}$$

$$\begin{aligned}
&\cdot\ f\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right)i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\\
&\geq D_{id\cdot f}(i,N,v_1,v_2)-f(1)\\
&+\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\\
&-\left(\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)f\left(\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right),
\end{aligned}\tag{4.68}$$

where $D_{id\cdot f}(i,N,v_1,v_2):=\sum_{i=1}^{N}\frac{1}{i^{v_1}H_{N,v_1}}f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)$.
If $t\mapsto tf(t)$ *is a concave function, then reverse inequality holds in* (4.68).

Proof. Inequality (4.67) leans on the proof of Theorem 4.1 wherein we insert for r_i and s_i expressions $\frac{1}{i^{v_1}H_{N,v_1}}$ and $\frac{1}{i^{v_2}H_{N,v_2}}$ by definition (4.13) of the Zipf law and $R_N=S_N=1$. Inequality (4.68) follows analogously after the proof of Theorem 4.2. Inequalities change their signs in the case of concavity of functions f or $t\mapsto tf(t)$ as a consequence of the Jensen inequality implicitly included. □

Monotonicity property of $D_f(i,N,v_1,v_2)$ and $D_{id\cdot f}(i,N,v_1,v_2)$ is presented in the sequel.

Corollary 4.18 *Let* f, v_1 *and* v_2 *be as in Corollary 4.17. Suppose* $\mathbf{t}=(t_1,\dots,t_N)$ *and* $\mathbf{u}=(u_1,\dots,u_N)$ *are positive real* $N-$*tuples such that* $T_N=\sum_{i=1}^{N}t_i$ *and* $U_N=\sum_{i=1}^{N}u_i$. *If* $\frac{1}{i^{v_2}H_{N,v_2}}\geq u_i$, *for* $i=1,\dots,N$ *then*

$$D_f(i,N,v_1,v_2)\geq f(1)+\sum_{i=1}^{N}u_i f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)-U_N f\left(\frac{1}{U_N}\sum_{i=1}^{N}u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right).\tag{4.69}$$

If $\frac{1}{i^{v_2}H_{N,v_2}}\leq t_i$, *for* $i=1,\dots,N$ *then*

$$D_f(i,N,v_1,v_2)\leq f(1)+\sum_{i=1}^{N}t_i f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)-T_N f\left(\frac{1}{T_N}\sum_{i=1}^{N}t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right).\tag{4.70}$$

If f *is a concave function, then reverse inequalities hold in* (4.69) *and* (4.70).

Suppose $t\mapsto tf(t)$ *is a convex function.*
If $\frac{1}{i^{v_2}H_{N,v_2}}\geq u_i$, *for* $i=1,\dots,N$ *then*

$$D_{id\cdot f}(i,N,v_1,v_2)\geq f(1)+\sum_{i=1}^{N}u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)$$

$$-\left(\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) f\left(\frac{1}{U_N}\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right), \tag{4.71}$$

If $\frac{1}{i^{v_2}H_{N,v_2}} \le t_i$, *for* $i = 1,\dots,N$ *then*

$$\begin{aligned} D_{id\cdot f}(i,N,v_1,v_2) \le f(1) &+ \sum_{i=1}^{N} t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) \\ &-\left(\sum_{i=1}^{N} t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) f\left(\frac{1}{T_N}\sum_{i=1}^{N} t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right), \end{aligned} \tag{4.72}$$

If $t \mapsto t f(t)$ *is a concave function, then reverse inequalities hold in* (4.71) *and* (4.72).

Proof. Inequalities (4.69), (4.70), (4.71) and (4.72) can be deduced from Corollary 4.17 or from Corollary 4.1 and Corollary 4.2 by analogous steps therein, if we observe the probability mass functions r_i and s_i as Zipf laws defined by (4.13). □

Remark 4.11 For $u_i = \min_{i=1,\dots,N}\left\{\frac{1}{i^{v_2}H_{N,v_2}}\right\} = \frac{1}{N^{v_2}H_{N,v_2}}$ and $t_i = \max_{i=1,\dots,N}\left\{\frac{1}{i^{v_2}H_{N,v_2}}\right\} = \frac{1}{H_{N,v_2}}$ inequalities (4.69) and (4.70) assume the form of the bounds that were obtained earlier in [11] :

$$f(1)+\frac{1}{H_{N,v_2}}\Lambda_3 \ge D_f(i,N,v_1,v_2) \ge f(1)+\frac{1}{N^{v_2}H_{N,v_2}}\Lambda_3, \tag{4.73}$$

where $\Lambda_3 = \sum_{i=1}^{N} f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) - Nf\left(\frac{1}{N}\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)$.
If we repeat the similar procedure with inequalities (4.71) and (4.72), we get the analogous bounds for $D_{id\cdot f}(i,N,v_1,v_2)$:

$$f(1)+\frac{1}{H_{N,v_2}}\widetilde{\Lambda}_3 \ge D_{id\cdot f}(i,N,v_1,v_2) \ge f(1)+\frac{1}{N^{v_2}H_{N,v_2}}\widetilde{\Lambda}_3, \tag{4.74}$$

where

$$\begin{aligned} \widetilde{\Lambda}_3 = &\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} f\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) \\ &-\left(\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) f\left(\frac{1}{N}\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right). \end{aligned} \tag{4.75}$$

In the sequel we present analogous results with the special choices on kernel function f. Although it is possible to deduce results concerning observed $f-$divergences for all three types of the Csiszár functional, that is for (4.46), (4.47) and (4.48), accompanied results for (4.48) will suffice in each implementation.

We start with Kullback-Leibler divergence (4.6) and superadditivity applied to both Zipf law $N-$tuples, *i.e.* those with $w_1 = w_2 = 0$.

Corollary 4.19 *Let $\overline{v} = (\overline{v}_1, \ldots, \overline{v}_N)$ be a positive real $N-$tuple such that $V_N = \sum_{i=1}^{N} \overline{v}_i$ and let $v_1, v_2 > 0$. If the logarithm base is greater than 1, then*

$$\begin{aligned}
&\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right) i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\log\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\\
&-\left(\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right) i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\\
&\cdot\ \log\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right) i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\\
&\geq KL(i,N,v_1,v_2)\\
&+\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\log\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\\
&-\left(\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\log\left(\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right).
\end{aligned} \tag{4.76}$$

If the logarithm base is less than 1, *then the inequality sign is reverse.*

Proof. Follows from (4.68) when observing function $t \mapsto \log t$ and its convexity (concavity) which depends on the logarithm base greater than 1 (less than 1.) □

Monotonicity can now be deduced directly from (4.76) and is given in the following corollary.

Corollary 4.20 *Suppose $\mathbf{t} = (t_1, \ldots, t_N)$ and $\mathbf{u} = (u_1, \ldots, u_N)$ are positive real $N-$tuples such that $T_N = \sum_{i=1}^{N} t_i$ and $U_N = \sum_{i=1}^{N} u_i$ and let $v_1, v_2 > 0$.*

If $\dfrac{1}{i^{v_2}H_{N,v_2}} \geq u_i$, for $i = 1, \ldots, N$ then

$$\begin{aligned}
KL(i,N,v_1,v_2) \geq &\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\log\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\\
&-\left(\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\log\left(\frac{1}{U_N}\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right),
\end{aligned} \tag{4.77}$$

where the logarithm base is greater than 1.
If $\frac{1}{i^{v_2}H_{N,v_2}} \leq t_i$, *for* $i = 1, \ldots, N$ *then*

$$KL(i,N,v_1,v_2) \leq \sum_{i=1}^{N} t_i i^{v_2-v_1} \frac{H_{N,v_2}}{H_{N,v_1}} \log\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) - \left(\sum_{i=1}^{N} t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\log\left(\frac{1}{T_N}\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right), \tag{4.78}$$

where the logarithm base is greater than 1.
If the logarithm base is less than 1, then reverse inequalities hold in (4.77) *and* (4.78).

Remark 4.12 If $u_i = \min_{i=1,\ldots,N}\left\{\frac{1}{i^{v_2}H_{N,v_2}}\right\} = \frac{1}{N^{v_2}H_{N,v_2}}$ and $t_i = \max_{i=1,\ldots,N}\left\{\frac{1}{i^{v_2}H_{N,v_2}}\right\} = \frac{1}{H_{N,v_2}}$ in inequalities (4.77) and (4.78), then the following bounds for the Kullback-Leibler divergence hold, as a special case of Corollary 4.20:

$$\frac{1}{H_{N,v_2}}\Lambda_{KL} \geq KL(i,N,v_1,v_2) \geq \frac{1}{N^{v_2}H_{N,v_2}}\Lambda_{KL}, \tag{4.79}$$

where

$$\Lambda_{KL} = \sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\log\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right) - \left(\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right)\log\left(\frac{1}{N}\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}\right). \tag{4.80}$$

Bounds (4.79) were obtained earlier in [11], due to a less general approach.

What follows are similar and concise results for the Hellinger distance, the Bhattacharyya coefficient, the chi-square divergence and the total variation distance.

Corollary 4.21 *Let* $\overline{v} = (\overline{v}_1, \ldots, \overline{v}_N)$ *be a positive real* $N-$*tuple such that* $V_N = \sum_{i=1}^{N}\overline{v}_i$ *and let* $v_1, v_2 > 0$. *Then*

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}} + \overline{v}_i\right)\left(\sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} - 1\right)^2 \\
&- \frac{1+\overline{V}_N}{2}\left(\sqrt{\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}} + \overline{v}_i\right)i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} - 1\right)^2 \\
&\geq h^2(i,N,v_1,v_2) + \frac{1}{2}\sum_{i=1}^{N}\overline{v}_i\left(\sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} - 1\right)^2
\end{aligned}$$

$$-\frac{\overline{V}_N}{2}\left(\sqrt{\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}-1\right)^2, \tag{4.81}$$

$$\begin{aligned}
&(1+\overline{V}_N)\sqrt{\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right)i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}\\
&-\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right)\sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}\\
&\geq B(i,N,v_1,v_2)+1-\sum_{i=1}^{N}\overline{v}_i\sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}\\
&+\overline{V}_N\sqrt{\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}},
\end{aligned} \tag{4.82}$$

$$\begin{aligned}
&\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right)\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right)^2\\
&-(1+\overline{V}_N)\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right)i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right)^2\\
&\geq \chi^2(i,N,v_1,v_2)+\sum_{i=1}^{N}\overline{v}_i\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right)^2\\
&-\overline{V}_N\left(\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right)^2,
\end{aligned} \tag{4.83}$$

$$\begin{aligned}
&\sum_{i=1}^{N}\left(\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right)\left|i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right|\\
&-(1+\overline{V}_N)\left(\frac{1}{1+\overline{V}_N}\sum_{i=1}^{N}\left|\frac{1}{i^{v_2}H_{N,v_2}}+\overline{v}_i\right)i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right|\\
&\geq V(i,N,v_1,v_2)+\sum_{i=1}^{N}\overline{v}_i\left|i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right|\\
&-\overline{V}_N\left|\frac{1}{\overline{V}_N}\sum_{i=1}^{N}\overline{v}_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right|.
\end{aligned} \tag{4.84}$$

Proof. Inequalities (4.81)-(4.84) follow from inequality (4.67) by inserting convex functions: $t\mapsto\frac{1}{2}\left(\sqrt{t}-1\right)^2$ for (4.81), $t\mapsto-\sqrt{t}$ for (4.82), $t\mapsto(t-1)^2$ for (4.83) and $t\mapsto|t-1|$ for (4.84). □

Again, superadditivity yields monotonicity which was independently given in [14] in a less general approach.

Corollary 4.22 *Let* **t** *and* **u** *be as in Corollary 4.18 and let* $v_1, v_2 > 0$.

If $\frac{1}{i^{v_2} H_{N,v_2}} \geq u_i$, *for* $i = 1, \ldots, N$ *then*

$$h^2(i,N,v_1,v_2) \geq \frac{1}{2}\sum_{i=1}^{N} u_i \left(\sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} - 1\right)^2 - \frac{U_N}{2}\left(\sqrt{\frac{1}{U_N}\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} - 1\right)^2, \tag{4.85}$$

$$B(i,N,v_1,v_2) \geq -1 - \sum_{i=1}^{N} u_i \sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} + U_N \sqrt{\frac{1}{U_N}\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}, \tag{4.86}$$

$$\chi^2(i,N,v_1,v_2) \geq \sum_{i=1}^{N} u_i \left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} - 1\right)^2 - U_N\left(\frac{1}{U_N}\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} - 1\right)^2, \tag{4.87}$$

$$V(i,N,v_1,v_2) \geq \sum_{i=1}^{N} u_i \left| i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} - 1\right| - U_N\left|\frac{1}{U_N}\sum_{i=1}^{N} u_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} - 1\right|. \tag{4.88}$$

If $\frac{1}{i^{v_2} H_{N,v_2}} \leq t_i$, *for* $i = 1, \ldots, N$ *then*

$$h^2(i,N,v_1,v_2) \leq \frac{1}{2}\sum_{i=1}^{N} t_i \left(\sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} - 1\right)^2 - \frac{T_N}{2}\left(\sqrt{\frac{1}{T_N}\sum_{i=1}^{N} t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} - 1\right)^2, \tag{4.89}$$

$$B(i,N,v_1,v_2) \leq -1 - \sum_{i=1}^{N} t_i \sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} + T_N \sqrt{\frac{1}{T_N}\sum_{i=1}^{N} t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}, \tag{4.90}$$

$$\chi^2(i,N,v_1,v_2) \leq \sum_{i=1}^{N} t_i \left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} - 1\right)^2 - T_N\left(\frac{1}{T_N}\sum_{i=1}^{N} t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} - 1\right)^2, \tag{4.91}$$

$$V(i,N,v_1,v_2) \leq \sum_{i=1}^{N} t_i \left| i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} - 1\right| - T_N\left|\frac{1}{T_N}\sum_{i=1}^{N} t_i i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}} - 1\right|. \tag{4.92}$$

Remark 4.13 If $u_i = \min_{i=1,\dots,N}\left\{\frac{1}{i^{v_2}H_{N,v_2}}\right\} = \frac{1}{N^{v_2}H_{N,v_2}}$ in inequalities (4.85)-(4.88) and $t_i = \max_{i=1,\dots,N}\left\{\frac{1}{i^{v_2}H_{N,v_2}}\right\} = \frac{1}{H_{N,v_2}}$ in inequalities (4.89)-(4.92), then the following bounds for the divergences hold, as special cases of Corollary 4.22.
Thus we have for Hellinger distance (4.7):

$$\frac{1}{2H_{N,v_2}}\Lambda_h \geq h^2(i,N,v_1,v_2) \geq \frac{1}{2N^{v_2}H_{N,v_2}}\Lambda_h, \tag{4.93}$$

where

$$\Lambda_h = \sum_{i=1}^{N}\left(\sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}-1\right)^2 - N\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}-1\right)^2; \tag{4.94}$$

for Bhattacharyya coefficient (4.8):

$$1-\frac{1}{N^{v_2}H_{N,v_2}}\Lambda_B \geq B(i,N,v_1,v_2) \geq 1-\frac{1}{H_{N,v_2}}\Lambda_B, \tag{4.95}$$

where

$$\Lambda_B = N\sqrt{\frac{1}{N}\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}} - \sum_{i=1}^{N}\sqrt{i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}}; \tag{4.96}$$

for chi-square divergence (4.9):

$$\frac{1}{H_{N,v_2}}\Lambda_{chi} \geq \chi^2(i,N,v_1,v_2) \geq \frac{1}{N^{v_2}H_{N,v_2}}\Lambda_{chi}, \tag{4.97}$$

where

$$\Lambda_{chi} = \sum_{i=1}^{N}\left(i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right)^2 - N\left(\frac{1}{N}\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right)^2 \tag{4.98}$$

and for total variation distance (4.10):

$$\frac{1}{H_{N,v_2}}\Lambda_V \geq V(i,N,v_1,v_2) \geq \frac{1}{N^{v_2}H_{N,v_2}}\Lambda_V, \tag{4.99}$$

where

$$\Lambda_V = \sum_{i=1}^{N}\left|i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right| - N\left|\frac{1}{N}\sum_{i=1}^{N} i^{v_2-v_1}\frac{H_{N,v_2}}{H_{N,v_1}}-1\right|. \tag{4.100}$$

Bounds (4.93), (4.95), (4.97) and (4.99) were also obtained earlier in [11], due to a less general approach.

Bibliography

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

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

[3] S. S. Dragomir, *Some inequalities for the Csiszár $\Phi-$ divergence*, RGMIA.

[4] S. S. Dragomir, J.E. Pečarić, L.E. Persson, *Properties of some functionals related to Jensen's inequality*, Acta Math. Hungar. **70** 1–2 (1996), 129–143.

[5] L. Egghe, R. Rousseau, *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Scince*, Elsevier Science Publishers, New York, (1990).

[6] T. van Erven, P. Harremoës, *Rényi Divergence and Kullback-Leibler Divergence*, Journal of Latex Class Files, **6** 1 (2007); arXiv: 1206.2459v2 [cs.IT]24 April 2014.

[7] M. Krnić, N. Lovričević, J. Pečarić, *Jessen's functional, its properties and applications*, An. Stiint. Univ. Ovidius Constanta Ser. Mat., Vol 20(1), (2012), 225–248.

[8] M. Krnić, N. Lovričević, J. Pečarić, *On McShane's functional's properties and its applications*, Period. Math. Hungar., Vol. 66 (2), (2013), 159–180.

[9] M. Krnić, N. Lovričević, J. Pečarić, *Superadditivity of the Levinson functional and applications*, Period. Math. Hungar., Vol. 71 (2), (2015), 166–178.

[10] M. Krnić, N. Lovričević, J. Pečarić, J. Perić, *Superadditivity and monotonicity of the Jensen-type functionals*, Element, Zagreb, 2015.

[11] S. Kullback, *Information Theory and Statistics*, J. Wiley, New York, 1959.

[12] S. Kullback, R. A. Leibler *On information and sufficiency*, Annals of Mathematical Statistics, **22** 1 (1951), 79–86.

[13] N. Lovričević, Đ. Pečarić, J. Pečarić *Zipf-Mandelbrot law, $f-$divergences and the Jensen-type interpolating inequalities*, J. Inequal. Appl. (2018) 2018:36 https://doi.org/10.1186/s13660-018-1625-y

[14] N. Lovričević, Đ. Pečarić, J. Pečarić *Monotonicity of the Jensen functional for $f-$divergences via the Zipf-Mandelbrot law*, submitted.

[15] B. Manaris, D. Vaughan, C. S. Wagner, J. Romero, R. B. Davis, *Evolutionary Music and the Zipf-Mandelbrot Law: Developing Fitness Functions for Pleasant Music*, Proceedings of 1st European Workshop on Evolutionary Music and Art (EvoMUSART2003), 522–534.

[16] B. Mandelbrot, *An informational theory of the statistical structure of language*, In Jackson, W., editor, Communication Theory, New York, Academic Press, 1953., 486–502.

[17] B. Mandelbrot, *Information Theory and Psycholinguistics*, in Wolman, B.B., and Nagel, E. (eds.), *Scientific psychology*, Basic Books, 1965.

[18] D. Yu. Manin, *Mandelbrot's Model for Zipf's Law: Can Mandelbrot's Model Explain Zipf's Law for Language?*, Journal of Quantitative Linguistics, **16** 3 (2009), 274–285.

[19] D. S. Mitrinović, J.E. Pečarić, A.M. Fink, *Classical and New Inequalities in Analysis*, Kluwer Academic Publishers, Dordrecht/Boston/London, 1993.

[20] M. A. Montemurro, *Beyond the Zipf-Mandelbrot law in quantitative linguistics*, URL:arxiv:condmat/0104066v2, (2001).

[21] D. Mouillot, A. Lepretre, *Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams(RFD), to assess changes in community diversity*, Environmental Monitoring and Assessment, Springer **63** (2) (2000), 279–295.

[22] Z. K. Silagadze, *Citations and the Zipf-Mandelbrot Law*, Complex Systems, **11** (1997), 487–499.

Chapter 5

Inequalities for Shannon and Zipf-Mandelbrot entropies by using Jensen's type inequalities

M. Adil Khan, Đilda Pečarić and Josip Pečarić

Abstract. Shannon and Zipf-Mandelbrot entropies have many applications in many applied sciences for example in Information Theory, Biology and Economics etc. In this paper we consider several Jensen type discrete as well as integral inequalities and obtain different bounds for Shannon and Zipf-Mandelbrot entropies. We also focus to investigate bounds for Csiszar divergence as well as hybrid Zipf-mandelbrot entropies.

2000 Mathematics Subject Classification. 26D15, 94A17, 94A15.
Key words and phrases. Shannon entropy, Zipf-Mandelbrot entropy, convex function, Jensen's inequality.

5.1 Introduction

Several functionals have been proposed in the literature as measures of information and each definition enjoys certain axiomatic and/or heuristic properties. A convenient way to differentiate among the various measures of information is to classify them in three categories: parametric, non-parametric and entropy-type measures of information. Parametric measures of information measure the amount of information supplied by the data about an unknown parameter θ and are functions of θ. In this case the best known measure is Fisher's measure of information [13]. Non-parametric measures express the amount of information supplied by the data for discriminating in favor of a distribution ϕ_1 against another ϕ_2 or measure the distance or affinity between ϕ_1 and ϕ_2. The best known measure of this type is the Kullback-Leibler measure[16]. Measures of entropy express the amount of information contained in a distribution, that is, the amount of uncertainty concerning the outcome of an experiment. The classical measures of this type are Shannon's [17] and Renyi's [29]. In this paper our focus will be on Shanon's entropy and Zipf-Mandelbrot entropy.

The concept of Shannon's entropy [17] is the central role of information theory sometimes referred as measure of uncertainty. The entropy of a random variable is defined in terms of its probability distribution and can be shown to be a good measure of randomness or uncertainty. Shannon entropy allows to estimate the average minimum number of bits needed to encode a string of symbols based on the alphabet size and the frequency of the symbols. The formula for Shannon entropy is given by

$$S(\mathbf{p}) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \tag{5.1}$$

where $p_{i's}$ are positive real numbers with $\sum_{i=1}^{n} p_i = 1$.

Zipf's law is one of the fundamental law in information science and it is very often used in linguistics. George Zipf's in 1932 found that we can count how many times each word appears in the text. So if we rank (r) word according to the frequency of word occurrence (f), then the product of these two numbers is a constant $(C) : C = r.f$.

Apart from the use of this law in information science and linguistics, Zipf's law is used in city populations, solar flare intensity, website traffic, earth quack magnitude and the size of moon craters etc. This distribution in economics is known as Pareto's law which analyze the distribution of the wealthiest members of the community [12, p. 125]. These two laws are the same in the mathematical sense, but they are applied in a different context [20, p. 294].

Benoit Mandelbrot in 1966 [18] gave a generalization of Zipf's law, known as Zipf-Mandelbrot law. Which gave improvement in account for the low-rank words in corpus where $k < 100$ [26]: $f(k) = \frac{c}{(k+q)^s}$, when we put $q = 0$; we get Zipf's law. Applications of Zipf-Mandelbrot law can be found in linguistics [26, 27], information sciences [28] and also mostly applicable in ecological field studies [29]. The formula for Zipf-Mandelbrot entropy is given by

$$Z(H,q,s) = \frac{s}{H_{n,q,s}} \sum_{k=1}^{n} \frac{\log(k+q)}{(k+q)^s} + \log H_{n,q,s}, \tag{5.2}$$

where $n \in \{1,2,3,...\}$, $q \geq 0$, $s > 0$, $k \in \{1,2,...,n\}$, $H_{n,q,s} = \sum_{i=1}^{n} \frac{1}{(i+q)^s}$ and the Zipf-Mandelbrot law(probability mass function) is given by:

$$f(k,n,q,s) = \frac{1/(k+q)^s}{H_{n,q,s}}. \tag{5.3}$$

Different scientific disciplines have different interpretation of the law. At this point, we will give interpretation from ecology [27]: parameters in (5.1) can be interpreted in the following way: n is the number of species present and the parameters, q represents the diversity of the environment and s the predictability of the ecosystem, i.e. the average probability of the appearance of species.

Further generalization of Zipf-Mandelbrot entropy is Hybrid Zipf-Mandelbrot entropy [15] which is given by

$$H(\Phi^*,q,s) := \frac{1}{\Phi^*(s,q,w)} \sum_{i=1}^{n} \frac{w^i}{(i+q)^s} \log\left(\frac{(i+q)^s}{w^i}\right) + \log \Phi^*(s,q,w)$$

where $\Phi^*(s,q,w) = \sum_{i=1}^{n} \frac{w^i}{(k+q)^s}$ and Hybrid Zipf-Mandelbrot law is given by

$$g(k,n,q,s) = \frac{w^k}{\Phi^*(s,q,w)(k+q)^s}.$$

In [15] the authors used maximum entropy approach [31] and proved that maximum value of the Shannon entropy under some constraints is Zipf-Mandelbrot law. That is, if $I = \{1,2,..,n\}$ or $I = \mathbb{N}$ for a given $q \geq 0$, a probability distribution that maximizes Shannon entropy under constraints $\sum\limits_{i \in I} p_i = 1$, $\sum\limits_{i \in I} p_i \ln(i+q) = \chi$ is Zipf-Mandelbrot law. They also used the same technique and derived hybrid Zipf-Mandelbrot law but with one additional constraint. That is, if $I = \{1,2,..,n\}$ or $I = \mathbb{N}$ for a given $q \geq 0$, a probability distribution that maximizes Shannon entropy under constraints $\sum\limits_{i \in I} p_i = 1$, $\sum\limits_{i \in I} p_i \ln(i+q) = \chi$, $\sum\limits_{i \in I} ip_i = \mu$ is hybrid Zipf-Mandelbrot law.

5.2 Discrete Jensen's type inequalities

The following improvement of Jensen's inequality has given in [30].

Theorem 5.1 *Let $\phi : [a,b] \to \mathbb{R}$ be a convex function, $x_i \in [a,b]$, $p_i > 0$, $\sum\limits_{i=1}^{n} p_i = P_n$. Then*

$$\frac{1}{P_n}\sum_{i=1}^{n} p_i\phi(x_i) \quad - \quad \phi(\overline{x}) \quad \geq \quad \left|\frac{1}{P_n}\sum_{i=1}^{n} p_i\left|\phi(x_i)-\phi(\overline{x})\right| - \frac{|\phi'_+(\overline{x})|}{P_n}\sum_{i=1}^{n} p_i\left|x_i-\overline{x}\right|\right|. \quad (5.4)$$

where, $\overline{x} = \frac{1}{P_n}\sum\limits_{i=1}^{n} p_i x_i$.

The following improvement of Jensen's inequality for monotone convex function has given in [1].

Theorem 5.2 *Let $\phi : [a,b] \to \mathbb{R}$ be a monotone convex function, $x_i \in [a,b]$, $p_i > 0$, $\sum_{i=1}^{n} p_i = 1$. If $x_i \geq \overline{x}$ for $i \in I \subset \{1,2,...,n\}$, then*

$$\sum_{i=1}^{n} p_i\phi(x_i) - \phi\left(\sum_{i=1}^{n} p_i x_i\right) \geq \left|\sum_{i=1}^{n} p_i sgn(x_i-\overline{x})\left[\phi(x_i)-x_i\phi'_+(\overline{x})\right]\right.$$
$$\left. + \left[\phi(\overline{x})-\overline{x}\phi'_+(\overline{x})\right][1-2P_I]\right|, \quad (5.5)$$

where, $\overline{x} = \sum_{i=1}^{n} p_i x_i$.

M. Matić and J. Pečarić in [21] proved general inequalities from which one can obtain some companion inequalities to the Jensen inequality:

Theorem 5.3 *Let $\phi : I \to \mathbb{R}$ be differentiable convex function defined on I. If $x_i \in I, i = 1,2,..,n (n \geq 2)$ are arbitrary members and $p_i \geq 0$ $(i = 1,2,..,n)$ with $P_n = \sum_{i=1}^{n} p_i > 0$ and let $\overline{x} = \frac{1}{P_n}\sum_{i=1}^{n} p_i x_i, \overline{y} = \frac{1}{P_n}\sum_{i=1}^{n} p_i\phi(x_i)$. If $c,d \in I$ are arbitrary chosen numbers, then we have*

$$\phi(c) + (\overline{x}-c)\phi'(c) \leq \overline{y} \leq \phi(d) + \frac{1}{P_n}\sum_{i=1}^{n} p_i(x_i-d)\phi'(x_i). \quad (5.6)$$

Also, when ϕ is strictly convex, we have equality in the left inequality in (5.14) if and only if $x_i = c$ holds for all indices i with $p_i > 0$, while equality holds in the right inequality in (5.14) if and only if $x_i = d$ holds for all indices i with $p_i > 0$.

If we set $c = d = \overline{x}$ in (5.14), then we can obtain Jensen's inequality as well as companion inequality to the Jensen inequality:

Theorem 5.4 ([22]) *Let $\phi : I \to \mathbb{R}$ be differentiable convex function defined on I and let $x_i, p_i, P_n, \overline{x}$ and $\overline{y}$ be stated as in* Theorem 5.25. *Then the inequalities*

$$0 \leq \overline{y} - \phi(\overline{x}) \leq \frac{1}{P_n}\sum_{i=1}^{n} p_i\phi'(x_i)(x_i-\overline{x}) \quad (5.7)$$

hold.

In the case when ϕ is strictly convex, we have equalities in (5.15) *if and only if there is some $c \in I$ such that $x_i = c$ holds for all i with $p_i > 0$.*

Moreover, if $\sum_{i=1}^{n} p_i\phi'(x_i) \neq 0$ and $\overline{\overline{x}} = \frac{\sum_{i=1}^{n} p_i x_i \phi'(x_i)}{\sum_{i=1}^{n} p_i\phi'(x_i)} \in I$, then by setting $d = \overline{\overline{x}}$, we get Slater's inequality:

Theorem 5.5 ([24]) *Suppose that $\phi : I \to \mathbb{R}$ is convex function on I, for $x_1, x_2, .., x_n \in I$ and for $p_1, p_2, .., p_n \geq 0$ with $P_n = \sum_{i=1}^{n} p_i > 0$. Let $\sum_{i=1}^{n} p_i\phi'_+(x_i) \neq 0$, $\frac{\sum_{i=1}^{n} p_i\phi'_+(x_i)x_i}{\sum_{i=1}^{n} p_i\phi'_+(x_i)} \in I$, then the following inequality holds:*

$$\frac{1}{P_n}\sum_{i=1}^{n} p_i\phi(x_i) \leq \phi\left(\frac{\sum_{i=1}^{n} p_i\phi'_+(x_i)x_i}{\sum_{i=1}^{n} p_i\phi'_+(x_i)}\right). \tag{5.8}$$

The following refinement of (5.14) has been given in [21].

Theorem 5.6 *Let $\phi : I \to \mathbb{R}$ be strictly convex differentiable function defined on I and let $x_i, p_i, P_n, \overline{x}$ and $\overline{y}$ be stated as in* Theorem 5.25 *and $\overline{d} = (\phi')^{-1}\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i\phi'(x_i)\right)$, then the following inequalities hold*

$$\overline{y} \leq \phi(\overline{d}) + \frac{1}{P_n}\sum_{i=1}^{n} p_i\phi'(x_i)(x_i - \overline{d}) \leq \phi(d) + \frac{1}{P_n}\sum_{i=1}^{n} p_i(x_i - d)\phi'(x_i). \tag{5.9}$$

Remark 5.1 In [25] Dragomir has also proved Theorem 5.12 for the case $d = \overline{x}$ and $P_n = 1$.

The following improvement of the right side of the inequality (5.14) has been given in [2].

Theorem 5.7 *Let $\phi : I \to \mathbb{R}$ be a convex function, $x_i \in I$, $p_i \geq 0$ $(i = 1, .., n)$ such that $P_n = \sum_{i=1}^{n} p_i > 0$ and $\overline{y} := \frac{1}{P_n}\sum_{i=1}^{n} p_i\phi(x_i)$, then we have*

$$\begin{aligned} &\phi(d) - \overline{y} - \frac{1}{P_n}\sum_{i=1}^{n} p_i\phi'_+(x_i)(d - x_i) \\ &\geq \left|\frac{1}{P_n}\sum_{i=1}^{n} p_i\left|\phi(d) - \phi(x_i)\right| - \frac{1}{P_n}\sum_{i=1}^{n} p_i\left|\phi'_+(x_i)(d - x_i)\right|\right|. \end{aligned} \tag{5.10}$$

As a consequence of the above theorem, the following improvement of Slater's inequality has been obtained:

Corollary 5.1 ([2]) *Let $\phi : I \to \mathbb{R}$ be a convex function, $x_i \in I$, $p_i \geq 0$ $(i = 1, .., n)$ such that $P_n = \sum_{i=1}^{n} p_i > 0$ and $\overline{y} := \frac{1}{P_n}\sum_{i=1}^{n} p_i\phi(x_i)$. If $\sum_{i=1}^{n} p_i\phi'_+(x_i) \neq 0$ such that $\overline{\overline{x}} = \frac{\sum_{i=1}^{n} p_i\phi'_+(x_i)x_i}{\sum_{i=1}^{n} p_i\phi'_+(x_i)} \in I$, then*

$$\phi(\overline{\overline{x}}) - \overline{y} \geq \left|\frac{1}{P_n}\sum_{i=1}^{n} p_i\left|\phi(\overline{\overline{x}}) - \phi(x_i)\right| - \frac{1}{P_n}\sum_{i=1}^{n} p_i\left|\phi'_+(x_i)(\overline{\overline{x}} - x_i)\right|\right|. \tag{5.11}$$

As a another consequence of Theorem 5.13, the following improvement of the inequality (5.15) has been obtained:

Corollary 5.2 ([2]) *Let $\phi : I \to \mathbb{R}$ be a convex function, $x_i \in I$, $p_i \geq 0$ $(i = 1,..,n)$ such that $P_n = \sum_{i=1}^n p_i > 0$, $\overline{y} := \frac{1}{P_n}\sum_{i=1}^n p_i\phi(x_i)$ and $\overline{x} := \frac{1}{P_n}\sum_{i=1}^n p_i x_i$, then we have*

$$\begin{aligned}
&\phi(\overline{x}) - \overline{y} - \frac{1}{P_n}\sum_{i=1}^n p_i\phi'_+(x_i)(\overline{x} - x_i) \\
&\geq \left|\frac{1}{P_n}\sum_{i=1}^n p_i\left|\phi(\overline{x}) - \phi(x_i)\right| - \frac{1}{P_n}\sum_{i=1}^n p_i\left|\phi'_+(x_i)(\overline{x} - x_i)\right|\right|. \qquad (5.12)
\end{aligned}$$

The following improvement of the left side of (5.14) is given in [3].

Theorem 5.8 *Let $\phi : I \to \mathbb{R}$ be a convex function, $x_i \in I$, $p_i \geq 0$ $(i = 1,..,n)$ such that $\sum_{i=1}^n p_i = 1$ and $\overline{x} := \sum_{i=1}^n p_i x_i$, then we have*

$$\begin{aligned}
&\sum_{i=1}^n p_i\phi(x_i) - \phi(c) - \phi'(c)(\overline{x} - c) \\
&\geq \left|\sum_{i=1}^n p_i\left|\phi(x_i) - \phi(c)\right| - \sum_{i=1}^n p_i\left|\phi'(c)(x_i - c)\right|\right|. \qquad (5.13)
\end{aligned}$$

5.3 Integral Jensen's type inequalities

Let (Ω, A, μ) be a measure space and $w : \Omega \to \mathbb{R}$ be measurable function with $w(x) > 0$ for $x \in \Omega$. We consider $L_w(\mu) := \{f : \Omega \to \mathbb{R}$ such that $\int_\Omega wf d\mu < \infty\}$.

Throughout this section $\overline{w}, \overline{f}$ and $\overline{g}$ are defined as:

$$\overline{w} = \int_\Omega w d\mu, \overline{g} = \frac{1}{\overline{w}}\int_\Omega w\phi(f)d\mu, \overline{f} = \frac{1}{\overline{w}}\int_\Omega wf d\mu,$$

where ϕ is a function such that the domain of ϕ is equal to the range of f.

Now we recall some results while these results are given in [21, 22, 24, 2, 3] for $w(x) = 1, x \in \Omega$.

M. Matić and J. Pečarić in [21] proved the following integral Jensen's type inequalities from which one can obtain some companion inequalities to the Jensen inequality:

Theorem 5.9 *Let (Ω, A, μ) be a measure space with $0 < \mu(\Omega) < \infty$ and let $\phi : (a,b) \to \mathbb{R}$ be differentiable convex function defined on interval (a,b). If $w : \Omega \to \mathbb{R}^+, f : \Omega \to (a,b)$ are such that $f, \phi(f), \phi'(f)$ and $\phi'(f)f$ are in $L_w(\mu)$, then for any $d \in (a,b)$ one has*

$$\phi(c) + (\overline{f} - c)\phi'(c) \leq \overline{g} \leq \phi(d) + \frac{1}{\overline{w}}\int_\Omega w(f - d)\phi'(f)d\mu. \qquad (5.14)$$

The following converse of Jensen's inequality is the the integral analogue of Theorem 2.1 given in [22], which can also be directly obtain from (5.14).

Theorem 5.10 *Let all the assumptions of Theorem 5.9 are satisfied. Then the following inequalities hold*

$$0 \leq \overline{g} - \phi(\overline{f}) \leq \frac{1}{\overline{w}} \int_{\Omega} w\phi'(f)(f - \overline{f})d\mu. \tag{5.15}$$

The following Slater's inequality has been given in [24].

Theorem 5.11 *Let (Ω, A, μ) be a measure space with $0 < \mu(\Omega) < \infty$ and let $\phi : (a,b) \to \mathbb{R}$ be a convex function defined on the interval (a,b). If $w : \Omega \to \mathbb{R}^+, f : \Omega \to (a,b)$ are such that $\phi(f), \phi'_+(f)$ and $\phi'_+(f)f$ are all in $L_w(\mu)$ and $\int_\Omega w\phi'_+(f)d\mu \neq 0$, $\frac{\int_\Omega w\phi'_+(f)fd\mu}{\int_\Omega w\phi'_+(f)d\mu} \in (a,b)$. Then the following inequality holds.*

$$\frac{1}{\overline{w}} \int_{\Omega} w\phi(f)d\mu \leq \phi\left(\frac{\int_\Omega wf\phi'_+(f)d\mu}{\int_\Omega w\phi'_+(f)d\mu}\right). \tag{5.16}$$

Remark 5.2 Let ϕ, f and $\overline{f}$ be stated as in Theorem 5.25, $\int_\Omega w\phi'(f)d\mu \neq 0$ and let $\overline{\overline{f}} = \frac{\int_\Omega w\phi'(f)fd\mu}{\int_\Omega w\phi'(f)d\mu} \in (a,b)$. If we put $d = \overline{\overline{f}}$ in (5.14) we immediately obtain Slater's inequality (5.36). On the other hand, if we put $d = \overline{f}$ in (5.14) we immediately obtain (5.15).

The following refinement of (5.15) is also valid (see [21]).

Theorem 5.12 *Let (Ω, A, μ) be a measure space with $0 < \mu(\Omega) < \infty$ and let $\phi : (a,b) \to \mathbb{R}$ be a differentiable, strictly convex function on interval (a,b). If $w : \Omega \to \mathbb{R}^+, f : \Omega \to (a,b)$ are such that $f, \phi(f), \phi'(f)$ and $\phi'(f)f$ are all in $L_w(\mu)$. Then there is exactly one $\overline{d} \in (a,b)$ such that $\phi'(\overline{d}) = \frac{1}{\overline{w}} \int_\Omega w\phi'(f)d\mu$ and*

$$\overline{g} \leq \phi(\overline{d}) + \frac{1}{\overline{w}} \int_{\Omega} w\phi'(f)(f - \overline{d})d\mu, \tag{5.17}$$

$$\begin{aligned} 0 \leq \overline{g} - \phi(\overline{f}) &\leq \phi(\overline{d}) + \frac{1}{\overline{w}} \int_{\Omega} w\phi'(f)(f - \overline{d})d\mu - \phi(\overline{f}) \\ &\leq \frac{1}{\overline{w}} \int_{\Omega} w\phi'(f)(f - \overline{f})d\mu. \end{aligned} \tag{5.18}$$

The following improvement of the right side of the inequality (5.14) has been given in [2].

Theorem 5.13 *Let (Ω, A, μ) be a measure space with $0 < \mu(\Omega) < \infty$ and $\phi : (a,b) \to \mathbb{R}$ be a convex function. If $w : \Omega \to \mathbb{R}^+, f : \Omega \to (a,b)$ are such that $\phi(f), \phi'_+(f)$ and $\phi'_+(f)f$ are in $L_w(\mu)$, then for any $d \in (a,b)$ we have*

$$\begin{aligned} &\phi(d) - \overline{g} - \frac{1}{\overline{w}} \int_{\Omega} w\phi'_+(f)(d - f)d\mu \\ &\geq \left| \frac{1}{\overline{w}} \int_{\Omega} w\left|\phi(d) - \phi(f)\right|d\mu - \frac{1}{\overline{w}} \int_{\Omega} w\left|\phi'_+(f)(f - d)\right|d\mu \right|. \end{aligned} \tag{5.19}$$

As a consequence of the above theorem, the following improvement of Slater's inequality has been obtained:

Corollary 5.3 ([2]) *Let all the assumptions of Theorem 5.13 hold. Then*

$$\phi(\bar{\bar{f}}) - \frac{1}{\overline{w}}\int_\Omega w\phi(f)d\mu \geq \left|\frac{1}{\overline{w}}\int_\Omega w\left|\phi(\bar{\bar{f}}) - \phi(f)\right|d\mu - \frac{1}{\overline{w}}\int_\Omega w\left|\phi'_+(f)(f-\bar{\bar{f}})\right|d\mu\right| \tag{5.20}$$

holds, whenever $\int_\Omega w\phi'_+(f)d\mu \neq 0$ *and* $\bar{\bar{f}} = \frac{\int_\Omega w\phi'_+(f)fd\mu}{\int_\Omega w\phi'_+(f)d\mu} \in (a,b)$.

As a another consequence of Theorem 5.13, the following improvement of the inequality (5.15) has been obtained:

Corollary 5.4 ([2]) *Let all the assumptions of Theorem 5.13 are satisfied. Then*

$$\begin{aligned}&\phi(\bar{f}) - \bar{g} - \frac{1}{\overline{w}}\int_\Omega w\phi'_+(f)(\bar{f}-f)d\mu \\ &\geq \left|\frac{1}{\overline{w}}\int_\Omega w\left|\phi(\bar{f}) - \phi(f)\right|d\mu - \frac{1}{\overline{w}}\int_\Omega w\left|\phi'_+(f)(f-\bar{f})\right|d\mu\right|.\end{aligned} \tag{5.21}$$

The following improvement of the left side of (5.14) is given in [3].

Theorem 5.14 *Let all the assumptions of Theorem 5.13 are satisfied. Then for any* $c \in (a,b)$ *we have*

$$\begin{aligned}&\bar{g} - \phi(c) - \phi'_+(c)(\bar{f}-c) \\ &\geq \left|\frac{1}{\overline{w}}\int_\Omega \left|w\phi(f) - \phi(c)\right|d\mu - \frac{|\phi'_+(c)|}{\overline{w}}\int_\Omega w\left|f-c\right|d\mu\right|.\end{aligned} \tag{5.22}$$

5.4 Bounds for Shannon entropy

In the following theorem we present bound for Shannon entropy by using refinement of Jensen's inequality for convex functions.

Theorem 5.15 ([4]) *Let* p_i *and* q_i $(i = 1,2,\ldots,n)$ *be positive real numbers with* $\sum_{i=1}^n p_i = 1$ *and* $\sum_{i=1}^n q_i = \alpha$, *then*

$$-S(\boldsymbol{p}) - \sum_{i=1}^n p_i \log q_i + \log\alpha \geq \left|\sum_{i=1}^n p_i\left|\log\alpha - \log\frac{q_i}{p_i}\right| - \frac{1}{\alpha}\sum_{i=1}^n p_i\left|\frac{q_i}{p_i} - \alpha\right|\right|. \tag{5.23}$$

Proof. Let $\phi(x) = -\log x$ and $x_i = \frac{q_i}{p_i}$, as $p_i, q_i > 0$ and $\sum_{i=1}^n q_i = \alpha$, then from (5.5) we get (5.23). □

The following corollary is the special case of Theorem 5.15.

Corollary 5.5 ([4]) *Let p_i $(i=1,2,...,n)$ be positive real numbers with $\sum_{i=1}^n p_i = 1$, then*

$$-S(\boldsymbol{p})+\log n \geq \left|\sum_{i=1}^n p_i \left|\log n+\log p_i\right| - \frac{1}{n}\sum_{i=1}^n p_i\left|\frac{1}{p_i}-n\right|\right|. \tag{5.24}$$

Proof. By taking $q_i = 1$, $i = 1,2,..,n$, in (5.23), we get (5.24). □

In the following theorem we obtain another bound for Shannon entropy by using refinement of Jensen's inequality for convex functions.

Theorem 5.16 ([4]) *Let p_i and q_i $(i=1,2,...,n)$ be positive real numbers with $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n q_i = \alpha$, then*

$$\begin{aligned}&-S(\boldsymbol{p})-\sum_{i=1}^n p_i\log q_i+\log\alpha\\ &\qquad\geq \left|\sum_{i=1}^n q_i\left|\frac{p_i}{q_i}\log\frac{p_i}{q_i}-\frac{1}{\alpha}\log\frac{1}{\alpha}\right| - \left|1-\log\alpha\right|\sum_{i=1}^n q_i\left|\frac{p_i}{q_i}-\frac{1}{\alpha}\right|\right|. \end{aligned}\tag{5.25}$$

Proof. Using inequality (5.5) for $\phi(x) = x\log x$ and then replacing p_i by q_i and x_i by $\frac{p_i}{q_i}$ we get

$$\begin{aligned}&\frac{1}{\alpha}\sum_{i=1}^n p_i\log\frac{p_i}{q_i}-\frac{1}{\alpha}\log\frac{1}{\alpha}\\ &\qquad\geq \left|\frac{1}{\alpha}\sum_{i=1}^n q_i\left|\frac{p_i}{q_i}\log\frac{p_i}{q_i}-\frac{1}{\alpha}\log\frac{1}{\alpha}\right| - \frac{1}{\alpha}\left|1+\log\frac{1}{\alpha}\right|\sum_{i=1}^n q_i\left|\frac{p_i}{q_i}-\frac{1}{\alpha}\right|\right|. \end{aligned}\tag{5.26}$$

Now using Logarithmic rules and definition of Shannon entropy we get (5.25). □

The following corollary is the special case of the above Theorem.

Corollary 5.6 ([4]) *Let p_i $(i=1,2,...,n)$ be positive real numbers with $\sum_{i=1}^n p_i = 1$, then*

$$-S(\boldsymbol{p})+\log n \geq \left|\sum_{i=1}^n \left|p_i\log p_i+\frac{1}{n}\log n\right| - \left|1-\log n\right|\sum_{i=1}^n\left|p_i-\frac{1}{n}\right|\right|. \tag{5.27}$$

Proof. By taking $q_i = 1$, $i = 1,2,..,n$, in (5.25) we get (5.27). □

In the following theorem we obtain bound for Shannon entropy by using refinement of Jensen's inequality for monotone convex functions.

Theorem 5.17 ([4]) *Let p_i and q_i $(i=1,2,...,n)$ be positive real numbers with $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n q_i = \alpha$. If $I = \{i \in \{1,2,..,n\} : \frac{p_i}{q_i} \geq \alpha\}$, then*

$$-S(\boldsymbol{p})-\sum_{i=1}^n p_i\log q_i+\log\alpha \geq \left|\sum_{i=1}^n p_i sgn\left(\frac{q_i}{p_i}-\alpha\right)\left[\frac{q_i}{\alpha p_i}-\log\frac{q_i}{p_i}\right]+\left[1-\log\alpha\right]\left[1-2P_I\right]\right|. \tag{5.28}$$

Proof. Let $\phi = -\log x$ and $x_i = \frac{q_i}{p_i}$ and $q_i > 0$, $\sum_{i=1}^{n} q_i = \alpha$, then from (5.5) we have

$$\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} + \log\alpha \geq \left| \sum_{i=1}^{n} p_i sgn\left(\frac{q_i}{p_i} - \alpha\right)\left[\frac{q_i}{\alpha p_i} - \log\frac{q_i}{p_i}\right] + [1-\log\alpha][1-2P_I] \right|. \tag{5.29}$$

Now using Logarithmic rules and definition of Shannon entropy we get (5.28) □

The following corollary is the special case of the above Theorem.

Corollary 5.7 ([4]) *Let p_i $(i = 1,2,...,n)$ be positive real numbers with $\sum_{i=1}^{n} p_i = 1$, then*

$$-S(\boldsymbol{p}) + \log n \geq \left| \sum_{i=1}^{n} p_i sgn\left(\frac{1}{p_i} - n\right)\left[\frac{1}{np_i} + \log p_i\right] + [1-\log n][1-2P_I] \right|. \tag{5.30}$$

Proof. By taking $q_i = 1$, $i = 1,2,..,n$, in (5.28), we get (5.58). □

5.5 Results for Csiszar divergence

In this section we give some basic results for Csiszar divergence and present applications of the results given in Section 1 for Csiszar divergence.

Given a convex function $f : \mathbb{R}_+ \to \mathbb{R}_+$, the $f-$divergence functional

$$I_\phi(\mathbf{p},\mathbf{q}) := \sum_{i=1}^{n} q_i \phi\left(\frac{p_i}{q_i}\right), \tag{5.31}$$

where $\mathbf{p} = (p_1,...,p_n)$, $\mathbf{q} = (q_1,...,q_n)$ are positive sequences, was introduced by Csiszár in [7], as a generalized measure of information, a distance function on the set of probability distributions $\mathbb{P}^n$. As in [7], we interpret undefined expressions by

$$\phi(0) = \lim_{t\to 0^+} f(t), \quad 0\phi\left(\frac{0}{0}\right) = 0,$$

$$0\phi\left(\frac{a}{0}\right) = \lim_{q\to 0^+} q\phi\left(\frac{a}{q}\right), \quad a\lim_{q\to\infty} \frac{\phi(t)}{t}, \quad a > 0.$$

The following results were essentially given by Csiszár and Körner [8]:

(i) If ϕ is convex, then $I_\phi(\mathbf{p},\mathbf{q})$ is jointly convex in $\mathbf{p}$ and $\mathbf{q}$;

(ii) For every $\mathbf{p},\mathbf{q} \in \mathbb{R}_+^n$, we have

$$I_\phi(\mathbf{p},\mathbf{q}) \geq \sum_{j=1}^{n} q_j \phi\left(\frac{\sum_{j=1}^{n} p_j}{\sum_{j=1}^{n} q_j}\right). \tag{5.32}$$

If ϕ is strictly convex, equality holds in (5.32) if and only if

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \ldots = \frac{p_n}{q_n}.$$

If ϕ is normalized, i.e., $\phi(1) = 0$, then for every $\mathbf{p}, \mathbf{q} \in \mathbb{R}^n_+$ with $\sum_{i=1}^n p_i = \sum_{i=1}^n q_i$, we have the inequality

$$I_\phi(\mathbf{p}, \mathbf{q}) \geq 0. \tag{5.33}$$

In particular, if $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, then (5.33) holds. This is the well-known positivity property of the ϕ-divergence.

In [3] the authors gave application of Theorem 5.9 for Csiszar divegence for the functions defined on the linear space. Here we mention that result for the function defined on an interval which is in fact application of Theorem 5.9.

Theorem 5.18 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be a convex function, $\mathbf{p} = (p_1, \ldots, p_n), \boldsymbol{q} = (q_1, \ldots, q_n)$ be two positive real n-tuples and let $P_n = \sum_{i=1}^n p_i$ and $Q_n = \sum_{i=1}^n q_i$. If $c, d \in \mathbb{R}_+$ are arbitrary chosen numbers, then the following inequalities hold*

$$\phi(c) + \left(\frac{P_n}{Q_n} - c\right)\phi'(c) \leq \frac{1}{Q_n} I_\phi(\boldsymbol{p}, \boldsymbol{q}) \leq \phi(d) + \frac{1}{Q_n}\sum_{i=1}^n p_i \phi'\left(\frac{p_i}{q_i}\right) - \frac{d}{Q_n} I_{\phi'}(\boldsymbol{p}, \boldsymbol{q}) \tag{5.34}$$

Remark 5.3 In [3] the authors gave Theorem 5.18 for the case $P_n = Q_n = 1$.

The following application of Theorem 5.25 for Csiszar divegence has been given in [23]. This result can also be obtained by using (5.18) for $c = d = \frac{P_n}{Q_n}$.

Theorem 5.19 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be a convex function, $\mathbf{p} = (p_1, \ldots, p_n), \boldsymbol{q} = (q_1, \ldots, q_n)$ be two positive real n-tuples and let $P_n = \sum_{i=1}^n p_i$ and $Q_n = \sum_{i=1}^n q_i$. Then*

$$0 \leq I_\phi(\boldsymbol{p}, \boldsymbol{q}) - Q_n \phi\left(\frac{P_n}{Q_n}\right) \leq \frac{1}{Q_n} I_{\phi'}\left(\frac{\boldsymbol{p}^2}{\boldsymbol{q}}, \boldsymbol{p}\right) - \frac{P_n}{Q_n} I_{\phi'}(\boldsymbol{p}, \boldsymbol{q}), \tag{5.35}$$

where $\frac{\boldsymbol{p}^2}{\boldsymbol{q}} = (\frac{p_1^2}{q_1}, \ldots, \frac{p_n^2}{q_n})$.

In the following result is the application of Theorem 5.5 for Csiszar divegence.

Theorem 5.20 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be a convex function, $\mathbf{p} = (p_1, \ldots, p_n), \boldsymbol{q} = (q_1, \ldots, q_n)$ be two positive real n-tuples and let $Q_n = \sum_{i=1}^n q_i$, $\frac{\sum_{i=1}^n p_i \phi'_+(\frac{p_i}{q_i})}{\sum_{i=1}^n q_i \phi'_+(\frac{p_i}{q_i})} \in \mathbb{R}_+$ and $\sum_{i=1}^n q_i \phi'_+(\frac{p_i}{q_i}) \neq 0$. Then*

$$\frac{1}{Q_n} I_\phi(\boldsymbol{p}, \boldsymbol{q}) \leq \phi\left(\frac{\sum_{i=1}^n p_i \phi'_+(\frac{p_i}{q_i})}{\sum_{i=1}^n q_i \phi'_+(\frac{p_i}{q_i})}\right). \tag{5.36}$$

In the following theorem we give application of Theorem 5.12 for Csiszar divegence.

Theorem 5.21 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be strictly convex and differentiable function, $\mathbf{p} = (p_1, \dots, p_n)$, $\boldsymbol{q} = (q_1, \dots, q_n)$ be two positive real n-tuples and let $Q_n = \sum_{i=1}^n q_i$ and $\overline{d} = (\phi')^{-1}\left(\frac{1}{Q_n} I_{\phi'}(\boldsymbol{p}, \boldsymbol{q})\right)$. Then for any $d \in \mathbb{R}_+$ we have*

$$\begin{aligned} \frac{1}{Q_n} I_\phi(\boldsymbol{p}, \boldsymbol{q}) &\le \phi(\overline{d}) + \frac{1}{Q_n} \sum_{i=1}^n q_i \phi'(\frac{p_i}{q_i}) \left(\frac{p_i}{q_i} - \overline{d}\right) \\ &\le \phi(d) + \frac{1}{Q_n} \sum_{i=1}^n q_i \left(\frac{p_i}{q_i} - d\right) \phi'(\frac{p_i}{q_i}). \end{aligned} \tag{5.37}$$

In the following theorem we give application of Theorem 5.13 for Csiszar divegence.

Theorem 5.22 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be convex function, $\mathbf{p} = (p_1, \dots, p_n), \boldsymbol{q} = (q_1, \dots, q_n)$ be two positive real n-tuples and let $Q_n = \sum_{i=1}^n q_i$, then we have*

$$\begin{aligned} &\phi(d) - \frac{1}{Q_n} I_\phi(\boldsymbol{p}, \boldsymbol{q}) - \frac{1}{Q_n} \sum_{i=1}^n q_i \phi'_+(\frac{p_i}{q_i}) \left(d - \frac{p_i}{q_i}\right) \\ &\ge \left| \frac{1}{Q_n} \sum_{i=1}^n q_i \left|\phi(d) - \phi(\frac{p_i}{q_i})\right| - \frac{1}{Q_n} \sum_{i=1}^n q_i \left|\phi'_+(\frac{p_i}{q_i})(d - \frac{p_i}{q_i})\right| \right|. \end{aligned} \tag{5.38}$$

The following result is a consequence of Theorem 5.22 which is in fact application of Corollary 5.1.

Corollary 5.8 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be a convex function, $\mathbf{p} = (p_1, \dots, p_n), \boldsymbol{q} = (q_1, \dots, q_n)$ be two positive real n-tuples and let $Q_n = \sum_{i=1}^n q_i$, $\overline{\overline{x}} := \frac{\sum_{i=1}^n p_i \phi'_+(\frac{p_i}{q_i})}{\sum_{i=1}^n q_i \phi'_+(\frac{p_i}{q_i})} \in \mathbb{R}_+$ and $\sum_{i=1}^n q_i \phi'_+(\frac{p_i}{q_i}) \neq 0$. Then*

$$\phi(\overline{\overline{x}}) - \frac{1}{Q_n} I_\phi(\boldsymbol{p}, \boldsymbol{q}) \ge \left| \frac{1}{Q_n} \sum_{i=1}^n q_i \left|\phi(\overline{\overline{x}}) - \phi(\frac{p_i}{q_i})\right| - \frac{1}{Q_n} \sum_{i=1}^n q_i \left|\phi'_+(\frac{p_i}{q_i})(\overline{\overline{x}} - \frac{p_i}{q_i})\right| \right|. \tag{5.39}$$

Another consequence of Theorem 5.22 which is in fact application of Corollary 5.2.

Corollary 5.9 *Let $\phi : \mathbb{R}_+ \to \mathbb{R}$ be a convex function, $\mathbf{p} = (p_1, \dots, p_n), \boldsymbol{q} = (q_1, \dots, q_n)$ be two positive real n-tuples and let $Q_n = \sum_{i=1}^n q_i$, then we have*

$$\begin{aligned} &\phi\left(\frac{P_n}{Q_n}\right) - \frac{1}{Q_n} I_\phi(\boldsymbol{p}, \boldsymbol{q}) - \frac{1}{Q_n} \sum_{i=1}^n q_i \phi'_+(\frac{p_i}{q_i}) \left(\frac{P_n}{Q_n} - \frac{p_i}{q_i}\right) \\ &\ge \left| \frac{1}{Q_n} \sum_{i=1}^n q_i \left|\phi\left(\frac{P_n}{Q_n}\right) - \phi(\frac{p_i}{q_i})\right| - \frac{1}{Q_n} \sum_{i=1}^n q_i \left|\phi'_+(\frac{p_i}{q_i}) \left(\frac{P_n}{Q_n} - \frac{p_i}{q_i}\right)\right| \right|. \end{aligned} \tag{5.40}$$

The following result can be obtained by using Theorem 5.8.

Theorem 5.23 *Let $\phi : R_+ \to \mathbb{R}$ be differentiable convex function, $\mathbf{p} = (p_1,...,p_n)$, $\boldsymbol{q} = (q_1,...,q_n)$ be two positive real n-tuples and let $P_n = \sum_{i=1}^n p_i$ and $Q_n = \sum_{i=1}^n q_i$. If $c \in [0,\infty]$ then we have*

$$\begin{aligned}&\frac{1}{Q_n} I_\phi(\boldsymbol{p},\boldsymbol{q}) - \phi(c) - \left(\frac{P_n}{Q_n} - c\right)\phi'(c) \\ &\geq \left| \frac{1}{Q_n}\sum_{i=1}^n q_i \left|\phi\left(\frac{p_i}{q_i}\right) - \phi(c)\right| - \frac{|\phi'(c)|}{Q_n}\sum_{i=1}^n q_i \left|\frac{p_i}{q_i} - c\right| \right|. \end{aligned} \tag{5.41}$$

Remark 5.4 In [3] the authors gave Theorem 5.23 for probability distributions **p** and **q**.

Now we start to give integral version of the above results:

Let (Ω, A, μ) be a measure space and $p : \Omega \to \mathbb{R}$ be measurable function. We consider $S := \{q : \Omega \to \mathbb{R}^+$ such that $\overline{q} := \int_\Omega q d\mu < \infty\}$. Let $\phi : [0,\infty) \to \mathbb{R}$ be a convex function.

In 1963, I. Csiszar [8] introduced the concept of ϕ-divergence as follows.

Definition 5.1 *Let $p, q \in S$, then the ϕ-divergence functional is defined by*

$$C_\phi(p,q) := \int_\Omega q(t)\phi\left(\frac{p(t)}{q(t)}\right) d\mu. \tag{5.42}$$

Now we start to give application of Theorem 5.9 for Csiszar divergence.

Theorem 5.24 *Let $\phi : (0,\infty) \to \mathbb{R}$ be a differentiable convex function and $p, q \in S$ with $\overline{p} := \int_\Omega p(t) d\mu, \overline{q} := \int_\Omega q(t) d\mu$, then for any $c, d \in (0,\infty)$ we have*

$$\phi(c) + \left(\frac{\overline{p}}{\overline{q}} - c\right)\phi'(c) \leq \frac{1}{\overline{q}} C_\phi(p,q) \leq \phi(d) + \frac{1}{\overline{q}}\int_\Omega p(t)\phi'\left(\frac{p(t)}{q(t)}\right) d\mu - \frac{d}{\overline{q}} C_{\phi'}(p,q). \tag{5.43}$$

Proof. Using Theorem 5.9 for $w \to q$ and $f \to \frac{p}{q}$, we obtain (5.43). □

The following result is the application of Theorem 5.25.

Corollary 5.10 *Let all the assumptions of Theorem 5.24 are satisfied. Then the following inequalities hold*

$$0 \leq C_\phi(p,q) - \overline{q}\phi\left(\frac{\overline{p}}{\overline{q}}\right) \leq C_\phi(\frac{p^2}{q}, p) - \frac{\overline{p}}{\overline{q}} C_\phi(p,q). \tag{5.44}$$

Proof. Putting $c = d = \frac{\overline{p}}{\overline{q}}$ in (5.43) we obtain (5.44). □

The following result is the application of Theorem 5.11 for Csiszar divegence.

Theorem 5.25 *Let $\phi : (0,\infty) \to \mathbb{R}$ be a convex function and $p, q \in S$ with $\overline{q} := \int_\Omega q(t) d\mu$ and let $\int_\Omega q(t)\phi'_+\left(\frac{p(t)}{q(t)}\right) d\mu \neq 0$ and $\frac{\int_\Omega p(t)\phi'_+\left(\frac{p(t)}{q(t)}\right) d\mu}{\int_\Omega q(t)\phi'_+\left(\frac{p(t)}{q(t)}\right) d\mu} \in \mathbb{R}_+$. Then*

$$\frac{1}{\overline{q}} C_\phi(p,q) \leq \phi\left(\frac{\int_\Omega p(t)\phi'_+\left(\frac{p(t)}{q(t)}\right) d\mu}{\int_\Omega q(t)\phi'_+\left(\frac{p(t)}{q(t)}\right) d\mu}\right). \tag{5.45}$$

The following theorem is the application of Theorem 5.12.

Theorem 5.26 *Let $\phi : (0,\infty) \to \mathbb{R}$ be strictly convex and differentiable function, and $p,q \in S$ with $\overline{q} := \int_\Omega q(t)d\mu$. Then there is exactly one $\overline{d} \in (0,\infty)$ such that $\phi'(\overline{d}) = \frac{1}{\overline{q}}C_{\phi'}(p,q)$ and for any $d \in (0,\infty)$ we have*

$$\frac{1}{\overline{q}}C_\phi(p,q) \le \phi(\overline{d}) + \frac{1}{\overline{q}}\int_\Omega q(t)\phi'\left(\frac{p(t)}{q(t)}\right)\left(\frac{p(t)}{q(t)} - \overline{d}\right)d\mu$$
$$\le \phi(d) + \frac{1}{\overline{q}}\int_\Omega q(t)\phi'\left(\frac{p(t)}{q(t)} - d\right)\phi'\left(\frac{p(t)}{q(t)}\right). \tag{5.46}$$

The following theorem is the application of Theorem 5.13 for Csiszar divergence.

Theorem 5.27 *Let $\phi : (0,\infty) \to \mathbb{R}$ be a convex function and $p,q \in S$ with $\overline{p} := \int_\Omega p(t)d\mu$, $\overline{q} := \int_\Omega q(t)d\mu$, then for any $d \in (0,\infty)$ we have*

$$\phi(d) - \frac{1}{\overline{q}}C_\phi(p,q) - \frac{1}{\overline{q}}\int_\Omega q(t)\phi'\left(\frac{p(t)}{q(t)}\right)\left(d - \frac{p(t)}{q(t)}\right)d\mu$$
$$\ge \left|\frac{1}{\overline{q}}\int_\Omega q(t)\left|\phi(d) - \phi\left(\frac{p(t)}{q(t)}\right)\right|d\mu\right.$$
$$\left. - \frac{1}{\overline{q}}\int_\Omega q(t)\left|\phi'_+\left(\frac{p(t)}{q(t)}\right)\left(\frac{p(t)}{q(t)} - d\right)\right|d\mu\right|. \tag{5.47}$$

The following result is a consequence of Theorem 5.27 which is in fact application of Corollary 5.3.

Corollary 5.11 *Let $\phi : (0,\infty) \to \mathbb{R}$ be a convex function and $p,q \in S$ with $\overline{q} := \int_\Omega q(t)d\mu$ and let $\int_\Omega q(t)\phi'_+\left(\frac{p(t)}{q(t)}\right)d\mu \neq 0$ and $\overline{\overline{d}} := \frac{\int_\Omega p(t)\phi'_+\left(\frac{p(t)}{q(t)}\right)d\mu}{\int_\Omega q(t)\phi'_+\left(\frac{p(t)}{q(t)}\right)d\mu} \in \mathbb{R}_+$. Then*

$$\phi(\overline{\overline{d}}) - \frac{1}{\overline{q}}C_\phi(p,q) \ge \left|\frac{1}{\overline{q}}\int_\Omega q(t)\left|\phi(\overline{\overline{d}}) - \phi\left(\frac{p(t)}{q(t)}\right)\right|d\mu\right.$$
$$\left. - \frac{1}{\overline{q}}\int_\Omega q(t)\left|\phi'_+\left(\frac{p(t)}{q(t)}\right)\left(\frac{p(t)}{q(t)} - \overline{\overline{d}}\right)\right|d\mu\right|. \tag{5.48}$$

Another consequence of Theorem 5.27 which is in fact application of Corollary 5.4.

Corollary 5.12 *Let all the assumptions of Theorem 5.9 are satisfied. Then the following inequalities hold*

$$\phi\left(\frac{\overline{p}}{\overline{q}}\right) - \frac{1}{\overline{q}}C_\phi(p,q) - \frac{1}{\overline{q}}\int_\Omega q(t)\phi'_+\left(\frac{p(t)}{q(t)}\right)\left(\frac{\overline{p}}{\overline{q}} - \frac{p(t)}{q(t)}\right)d\mu$$
$$\ge \left|\frac{1}{\overline{q}}\int_\Omega q(t)\left|\phi\left(\frac{\overline{p}}{\overline{q}}\right) - \phi\left(\frac{p(t)}{q(t)}\right)\right|d\mu\right.$$
$$\left. - \frac{1}{\overline{q}}\int_\Omega q(t)\left|\phi'_+\left(\frac{p(t)}{q(t)}\right)\left(\frac{p(t)}{q(t)} - \frac{\overline{p}}{\overline{q}}\right)\right|d\mu.\right| \tag{5.49}$$

The following result can be obtained by using Theorem 5.14.

Theorem 5.28 *Let $\phi : (0,\infty) \to \mathbb{R}$ be a convex function and $p,q \in S$ with $\overline{p} := \int_\Omega p(t)d\mu$, $\overline{q} := \int_\Omega q(t)d\mu$, then for any $c \in (0,\infty)$, we have*

$$\begin{aligned}\frac{1}{\overline{q}}C_\phi(p,q) - \phi(c) - \phi'_+(c)\left(\frac{\overline{p}}{\overline{q}} - c\right)\\ \geq \left|\frac{1}{\overline{q}}\int_\Omega \left|q(t)\phi\left(\frac{p(t)}{q(t)}\right) - \phi(c)\right|d\mu - \frac{|\phi'_+(c)|}{\overline{q}}\int_\Omega q(t)\left|\frac{p(t)}{q(t)} - c\right|d\mu\right|. \qquad (5.50)\end{aligned}$$

5.6 Bounds for Zipf-Mandelbrot entropy

In the following theorem we use Zipf-Mandelbrot law and deduce bounds for Zipf-Mandelbrot entropy.

Theorem 5.29 ([4]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $s > 0$, $q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^n q_i = \alpha$, then*

$$\begin{aligned}-Z(H,q,s) - \sum_{i=1}^n \frac{\log q_i}{(i+q)^s H_{n,q,s}} + \log\alpha \geq \Bigg|\sum_{i=1}^n \frac{1}{(i+q)^s H_{n,q,s}}\left|\log\alpha - \log(q_i(i+q)^s H_{n,q,s})\right|\\ - \frac{1}{\alpha}\sum_{i=1}^n \frac{1}{(i+q)^s H_{n,q,s}}\left|q_i(i+q)^s H_{n,q,s} - \alpha\right|\Bigg|. \qquad (5.51)\end{aligned}$$

Proof. Let $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,...,n$, then

$$\begin{aligned}\sum_{i=1}^n p_i \log p_i &= \sum_{i=1}^n \frac{1}{(i+q)^s H_{n,q,s}} \log\frac{1}{(i+q)^s H_{n,q,s}}\\ &= -\sum_{i=1}^n \frac{1}{(i+q)^s H_{n,q,s}} \log((i+q)^s H_{n,q,s})\\ &= -\sum_{i=1}^n \frac{s}{(i+q)^s H_{n,q,s}} \log(i+q) - \sum_{i=1}^n \frac{\log H_{n,q,s}}{(i+q)^s H_{n,q,s}}\\ &= -\frac{s}{(i+q)^s H_{n,q,s}}\sum_{i=1}^n \frac{\log(i+q)}{(i+q)^s} - \frac{\log H_{n,q,s}}{H_{n,q,s}}\sum_{i=1}^n \frac{1}{(i+q)^s} = -Z(H,q,s).\end{aligned}$$

As $H_{n,q,s} = \sum_{i=1}^n \frac{1}{(i+q)^s}$, therefore $\sum_{i=1}^n \frac{1}{(i+q)^s H_{n,q,s}} = 1$. Hence using (5.23) for $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,..,n$, we obtain (5.51). □

The following corollary is the special case of above theorem.

Corollary 5.13 ([4]) *Let $n \in \{1,2,3,\dots\}$, $q \geq 0$, $s > 0$, then*

$$-Z(H,q,s)+\log n \geq \left| \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left|\log n - \log((i+q)^s H_{n,q,s})\right| - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left|(i+q)^s H_{n,q,s} - n\right| \right|. \quad (5.52)$$

Proof. By taking $q_i = 1$, $i = 1,2,..,n$, in (5.51), we get (5.52). □

In the following theorem we use two Zipf-Mandelbrot laws for different parameters and deduce bound for Zipf-Mandelbrot entropy.

Theorem 5.30 ([4]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then*

$$-Z(H,t_1,s_1)+\sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \geq \left| \sum_{i=1}^{n} \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \left|\log \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{(i+t_2)^{s_2} H_{n,t_2,s_2}}\right| - \sum_{i=1}^{n} \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \left|\frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{(i+t_2)^{s_2} H_{n,t_2,s_2}} - 1\right| \right|. \quad (5.53)$$

Proof. Let $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,\dots,n$, then as in the proof of Theorem 5.29 we have

$$\sum_{i=1}^{n} p_i \log p_i = \sum_{i=1}^{n} \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} = -Z(H,t_1,s_1).$$

$$\sum_{i=1}^{n} p_i \log q_i = \sum_{i=1}^{n} \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} = -\sum_{i=1}^{n} \frac{\log((i+t_2)^{s_1} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}}.$$

$$\text{Also } \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} = \alpha = 1 \text{ and } \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} = 1.$$

Therefore using (5.23) for $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,..,n$, we obtain (5.53). □

Theorem 5.31 ([4]) *Let $n \in \{1,2,3,\dots\}$, $q > 0$, $s > 0$, $q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} q_i = \alpha$, then*

$$\begin{aligned} -Z(H,q,s) &- \sum_{i=1}^{n} \frac{\log q_i}{(i+q)^s H_{n,q,s}} + \log \alpha \\ &\geq \left| \sum_{i=1}^{n} q_i \left| \frac{1}{q_i (i+q)^s H_{n,q,s}} \log(q_i (i+q)^s H_{n,q,s}) + \frac{1}{\alpha} \log \frac{1}{\alpha} \right| \right. \\ &- \left. |1-\log\alpha| \sum_{i=1}^{n} q_i \left| \frac{1}{q_i (i+q)^s H_{n,q,s}} - \frac{1}{\alpha} \right| \right|. \quad (5.54) \end{aligned}$$

Proof. As in the proof of Theorem 5.29, using (5.25) for $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,..,n$, we get (5.54). □

Corollary 5.14 ([4]) *Let* $n \in \{1,2,3,...\}$, $q > 0$, $s > 0$, *then*

$$\begin{aligned}
-Z(H,q,s) + \log n & \\
\geq & \left| \sum_{i=1}^{n} \left| \frac{1}{(i+q)^s H_{n,q,s}} \log((i+q)^s H_{n,q,s}) + \frac{1}{n} \log \frac{1}{n} \right| \right. \\
& \left. - \ |1 - \log n| \sum_{i=1}^{n} \left| \frac{1}{(i+q)^s H_{n,q,s}} - \frac{1}{\alpha} \right| \right|. \qquad (5.55)
\end{aligned}$$

Proof. By taking $q_i = 1$, $i = 1,2,..,n$, in (5.54), we get (5.55). □

Theorem 5.32 ([4]) *Let* $n \in \{1,2,3,...\}$, $t_1, t_2 \geq 0$ *and* $s_1, s_2 > 0$, *then*

$$\begin{aligned}
-Z(H,t_1,s_1) \ + & \sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \\
\geq & \left| \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \log \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \right| \right. \\
& \left. - \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - 1 \right| \right|. \qquad (5.56)
\end{aligned}$$

Proof. As in the proof of Theorem 5.30, using (5.25) for $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,..,n$, we get (5.56). □

Theorem 5.33 ([4]) *Let* $n \in \{1,2,3,...\}$, $q > 0$, $s > 0$, $q_i > 0$, $i = 1,2,..,n$ *with* $\sum_{i=1}^{n} q_i = \alpha$. *If* $I = \{i \in \{1,2,..,n\} : \frac{1}{q_i (i+q)^s H_{n,q,s}} \geq \alpha\}$, *then*

$$\begin{aligned}
& - Z(H,q,s) - \sum_{i=1}^{n} \frac{\log q_i}{(i+q)^s H_{n,q,s}} + \log \alpha \\
& \geq \left| \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} sgn\,(q_i (i+q)^s H_{n,q,s} - \alpha) \left[\frac{q_i (i+q)^s H_{n,q,s}}{\alpha} - \log q_i (i+q)^s H_{n,q,s} \right] \right. \\
& \qquad \left. + [1 - \log \alpha] \left[1 - 2 \sum_{i \in I} \frac{1}{(i+q)^s H_{n,q,s}} \right] \right|. \qquad (5.57)
\end{aligned}$$

Proof. Using (5.28) for $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$, we get (5.57). □

Corollary 5.15 ([4]) *Let $n \in \{1,2,3,\dots\}$, $q>0$, $s>0$, $k \in \{1,2,\dots,n\}$, then*

$$
\begin{aligned}
&-Z(H,q,s)+\log n \\
&\geq \left| \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} sgn\left((i+q)^s H_{n,q,s}-n\right)\left[\frac{(i+q)^s H_{n,q,s}}{n} - \log(i+q)^s H_{n,q,s}\right] \right. \\
&\qquad \left. +[1-\log n]\left[1-2\sum_{i=1}^{k}\frac{1}{(i+q)^s H_{n,q,s}}\right]\right|. \qquad (5.58)
\end{aligned}
$$

Proof. Since $\frac{1}{(i+q)^s H_{n,q,s}}$ is decreasing sequence over $i=1,2,..,n$. Therefore there exists $k \in \{1,2,..,n\}$ such that $\frac{1}{(i+q)^s H_{n,q,s}} \geq n$ for all $i \in \{1,2,..,k\}$. Now By taking $q_i = 1$, $i=1,2,..,n$, in (5.57), we get (5.58). □

Theorem 5.34 ([4]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$ and $s_1,s_2>0$. If $I=\{i \in \{1,2,..,n\} : \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \geq 1\}$, then*

$$
\begin{aligned}
&-Z(H,t_1,s_1)+\sum_{i=1}^{n}\frac{\log((i+t_1)^{s_1}H_{n,t_1,s_1})}{(i+t_1)^{s_1}H_{n,t_1,s_1}} \\
&\geq \left|\sum_{i=1}^{n}\frac{sgn\left(\frac{(i+t_1)^{s_1}H_{n,t_1,s_1}}{(i+t_2)^{s_2}H_{n,t_2,s_2}}-1\right)}{(i+t_1)^{s_1}H_{n,t_1,s_1}}\left[\frac{(i+t_1)^{s_1}H_{n,t_1,s_1}}{(i+t_2)^{s_2}H_{n,t_2,s_2}}-\log\left(\frac{(i+t_1)^{s_1}H_{n,t_1,s_1}}{(i+t_2)^{s_2}H_{n,t_2,s_2}}\right)\right]\right. \\
&\qquad \left. +\left[1-2\sum_{i\in I}\frac{1}{(i+t_1)^{s_1}H_{n,t_1,s_1}}\right]\right|. \qquad (5.59)
\end{aligned}
$$

Proof. Using (5.28) for $p_i = \frac{1}{(i+t_1)^{s_1}H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2}H_{n,t_2,s_2}}$, we get (5.59). □

We start to give first general inequalities for Zipf-Mandelbrot entropy which contain two arbitrary positive real numbers.

Theorem 5.35 ([5]) *Let $n \in \{1,2,3,\dots\}$, $q \geq 0$, $s, p_i > 0$, $i=1,2,..,n$ with $\sum_{i=1}^{n} p_i = P_n$, then for any $c,d \in \mathbb{R}_+$, we have*

$$
\log c+(P_n-c)\frac{1}{c} \geq \sum_{i=1}^{n}\frac{\log p_i}{(i+q)^s H_{n,q,s}}+Z(H,q,s) \geq \log d+1-d\sum_{i=1}^{n}\frac{1}{p_i(i+q)^{2s}H_{n,q,s}^2}. \qquad (5.60)
$$

Proof. Using $\phi(x) = -\log x$ in (5.34) we obtain

$$
-Q_n \log c-(P_n-cQ_n)\frac{1}{c} \leq \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right) \leq -Q_n\log d - Q_n + d\sum_{i=1}^{n}\frac{q_i^2}{p_i}. \qquad (5.61)
$$

Let $q_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i=1,2,\dots,n$, then $Q_n = \sum_{i=1}^{n} q_i = 1$ and

$$
\sum_{i=1}^{n} q_i \log q_i = \sum_{i=1}^{n}\frac{1}{(i+q)^s H_{n,q,s}}\log\frac{1}{(i+q)^s H_{n,q,s}}
$$

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

and

$$
\sum_{i=1}^{n} q_i \log p_i = \sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}}
$$

Therefore (5.61) implies that

$$
-\log c - (P_n - c)\frac{1}{c} \leq -\sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} - Z(H,q,s)
$$

$$
\leq -\log d - 1 + d\sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}, \tag{5.62}
$$

which is equivalent to (5.60). □

The following corollaries are consequences of the above theorem.

Corollary 5.16 ([5]) *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s, p_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} p_i = P_n$, then*

$$
0 \geq \sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} + Z(H,q,s) - \log P_n \geq 1 - P_n \sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}. \tag{5.63}
$$

Proof. Take $c = d = P_n$ in (5.60), we get (5.63). □

The following consequence of Theorem 5.50 is in fact the application of Slater's inequality.

Corollary 5.17 ([5]) *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $s, p_i > 0$, $i = 1,2,..,n$, then*

$$
\sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} + Z(H,q,s) \geq \log\left(\frac{1}{\sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}}\right). \tag{5.64}
$$

Proof. By setting $d = \frac{1}{\sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}}$ in the right inequality of (5.60), we get (5.64). □

In the following theorem we give general inequalities for Zipf-Mandelbrot entropies corresponding to different parameters.

Theorem 5.36 ([5]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then for any $c,d \in \mathbb{R}_+$, we have*

$$\log c + (1-c)\frac{1}{c} \geq Z(H,t_2,s_2) - \sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$$
$$\geq \log d + 1 - d\sum_{i=1}^{n} \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{((i+t_2)^{s_2} H_{n,t_2,s_2})^2}. \tag{5.65}$$

Proof. Let $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,\dots,n$, then

$$\sum_{i=1}^{n} q_i \log q_i = \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \log \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} = -Z(H,t_2,s_2),$$

$$\sum_{i=1}^{n} q_i \log p_i = \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \log \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} = -\sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}}.$$

$$\text{Also } Q_n = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} = 1 \text{ and } P_n = \sum_{i=1}^{n} \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} = 1$$

Therefore using (5.61) for $p_i = \frac{1}{(i+q)^{s} H_{n,q,s}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,..,n$, we obtain (5.65). □

The following corollaries are the consequence of Theorem 5.50.

Corollary 5.18 ([5]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then*

$$0 \geq Z(H,t_2,s_2) - \sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \geq 1 - \sum_{i=1}^{n} \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{((i+t_2)^{s_2} H_{n,t_2,s_2})^2}. \tag{5.66}$$

Proof. Take $c = d = 1$ in (5.65), we get (5.66). □

The following corollary is the application of Theorem 5.50 is in fact the application of Slater's inequality for Zipf-Mandelbrot entropies corresponding to different parameters.

Corollary 5.19 ([5]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then*

$$\sum_{i=1}^{n} \frac{\log((i+t_1)^{s_2} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}} - Z(H,t_2,s_2) \leq \log\left(\sum_{i=1}^{n} \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{((i+t_2)^{s_2} H_{n,t_2,s_2})^2}\right). \tag{5.67}$$

Proof. Take $d = \frac{1}{\sum_{i=1}^{n} \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{((i+t_2)^{s_2} H_{n,t_2,s_2})^2}}$ in the right inequality of (5.65), we get (5.103). □

The following result for Zipf-Mandelbrot entropy has been obtained by using the right inequality in (5.37).

Theorem 5.37 ([5]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $s, p_i > 0$, $i = 1,2,..,n$, then for any $d \in \mathbb{R}_+$ we have*

$$\sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} + Z(H,q,s) \geq \log\left(\frac{1}{\sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}}\right)$$
$$\geq \log d + 1 - d \sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}. \tag{5.68}$$

Proof. By setting $\phi(x) = -\log x$ in (5.37) we have

$$\frac{-1}{Q_n} \sum_{i=1}^{n} q_i \log \frac{p_i}{q_i} \leq -\log \frac{Q_n}{\sum_{i=1}^{n} \frac{q_i^2}{p_i}} \leq -\log d - \frac{1}{Q_n} \sum_{i=1}^{n} \frac{q_i^2}{p_i}\left(\frac{p_i}{q_i} - d\right). \tag{5.69}$$

Now putting $q_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,...,n$, in (5.105) we deduce

$$-\sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} - Z(H,q,s) \leq -\log\left(\frac{1}{\sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}}\right)$$
$$\leq -\log d - 1 + d \sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}.$$

which is equivalent to (5.104). □

In the following theorem we have presented result for two Zipf-Mandelbrot entropies corresponding to different parameters.

Theorem 5.38 ([5]) *Let $n \in \{1,2,3,...\}$, $t_1, t_2 \geq 0$, $s_1, s_2 > 0$, then for any $d \in \mathbb{R}_+$, we have*

$$\sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}} - Z(H,t_2,s_2) \leq \log \sum_{i=1}^{n} \frac{((i+t_2)^{s_2} H_{n,t_2,s_2})^2}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$$
$$\leq -\log d - 1 + \frac{d(i+t_1)^{s_1} H_{n,t_1,s_1}}{((i+t_2)^{s_2} H_{n,t_2,s_2})^2}. \tag{5.70}$$

Proof. As in the proof of Theorem 5.37, putting $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,...,n$, in (5.105), we get (5.106). □

The following application of the inequality (5.38) for Zipf-Mandelbrot entropy holds.

Theorem 5.39 ([5]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $s, p_i > 0$, $i = 1,2,..,n$, then for any $d \in \mathbb{R}_+$, we have*

$$-\log d + \sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} + Z(H,q,s) + \sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2} (d - p_i (i+q)^s H_{n,q,s})$$

$$\geq \left| \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left| \log d - \log p_i (i+q)^s H_{n,q,s} \right| \right.$$
$$\left. - \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left| \frac{d}{p_i (i+q)^s H_{n,q,s}} - 1 \right| \right|. \tag{5.71}$$

Proof. By using (5.38) for $\phi(x) = -\log x$, we have

$$-\log d + \sum_{i=1}^{n} q_i \log \frac{p_i}{q_i} + \frac{1}{Q_n} \sum_{i=1}^{n} \frac{q_i^2}{p_i} \left(d - \frac{p_i}{q_i} \right)$$
$$\geq \left| \frac{1}{Q_n} \sum_{i=1}^{n} q_i \left| \log d - \log \frac{p_i}{q_i} \right| - \frac{1}{Q_n} \sum_{i=1}^{n} q_i \left| \frac{q_i}{p_i} (d - \frac{p_i}{q_i}) \right| \right|. \tag{5.72}$$

Now putting $q_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1, 2, \ldots, n$, in (5.72) we obtain (5.71). □

The following consequence of the above theorem is in fact the refinement of inequality (5.64).

Corollary 5.20 ([5]) *Let $n \in \{1, 2, 3, \ldots\}$, $q \geq 0$, $s, p_i > 0$, $i = 1, 2, .., n$ and $\tilde{x} := \frac{1}{\sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2}}$. Then*

$$-\log \tilde{x} + \sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} + Z(H, q, s)$$
$$\geq \left| \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left| \log \tilde{x} - \log p_i (i+q)^s H_{n,q,s} \right| \right.$$
$$\left. - \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left| \frac{\tilde{x}}{p_i (i+q)^s H_{n,q,s}} - 1 \right| \right|. \tag{5.73}$$

Proof. By setting $d = \tilde{x}$ in (5.71) we get (5.73). □

Corollary 5.21 ([5]) *Let $n \in \{1, 2, 3, \ldots\}$, $q \geq 0$, $s, p_i > 0$, $i = 1, 2, .., n$, then we have*

$$-\log P_n + \sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} + Z(H, q, s) + \sum_{i=1}^{n} \frac{1}{p_i (i+q)^{2s} H_{n,q,s}^2} (P_n - p_i (i+q)^s H_{n,q,s})$$
$$\geq \left| \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left| \log P_n - \log p_i (i+q)^s H_{n,q,s} \right| \right.$$
$$\left. - \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left| \frac{P_n}{p_i (i+q)^s H_{n,q,s}} - 1 \right| \right|. \tag{5.74}$$

Proof. By setting $d = P_n$ in (5.71) we get (5.74). □

In the following theorem we have given refinement of the right inequality in (5.65).

Theorem 5.40 ([5]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then for any $d \in \mathbb{R}_+$, we have*

$$
\begin{aligned}
-\log d + Z(H,t_2,s_2) - \sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}} + \sum_{i=1}^{n} \frac{d(i+t_1)^{s_1} H_{n,t_1,s_1}}{((i+t_2)^{s_2} H_{n,t_2,s_2})^2} - 1 \\
\geq \left| \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \log d - \log \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \right| \right. \\
\left. - \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \frac{d(i+t_1)^{s_1} H_{n,t_1,s_1}}{(i+t_2)^{s_2} H_{n,t_2,s_2}} - 1 \right| \right|. \qquad (5.75)
\end{aligned}
$$

Proof. Putting $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,\dots,n$, in (5.72), we obtain (5.75). □

The following consequence of the above theorem is in fact refinement of Slater's inequality for Zipf-Mandelbrot entropies corresponding to different parameters.

Corollary 5.22 ([5]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$ and $\tilde{x} = \frac{1}{\sum_{i=1}^{n} \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{((i+t_2)^{s_2} H_{n,t_2,s_2})^2}}$, then*

$$
\begin{aligned}
-\log \tilde{x} + Z(H,t_2,s_2) - \sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \\
\geq \left| \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \log \tilde{x} - \log \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \right| \right. \\
\left. - \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \frac{\tilde{x}(i+t_1)^{s_1} H_{n,t_1,s_1}}{(i+t_2)^{s_2} H_{n,t_2,s_2}} - 1 \right| \right|. \qquad (5.76)
\end{aligned}
$$

Proof. By setting $d = \frac{1}{\sum_{i=1}^{n} \frac{(i+t_1)^{s_1} H_{n,t_1,s_1}}{((i+t_2)^{s_2} H_{n,t_2,s_2})^2}}$ in (5.75), we obtain (5.76). □

The following theorem is the application of Theorem 5.23 for Zipf-Mandelbrot entropy.

Theorem 5.41 ([5]) *Let $n \in \{1,2,3,\dots\}$, $q \geq 0$, $s, p_i > 0$, $i = 1,2,..,n$, then for any $c \in \mathbb{R}_+$, we have*

$$
\begin{aligned}
-\sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} - Z(H,q,s) + \log c + (P_n - c)\frac{1}{c} \\
\geq \left| \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left| \log(p_i (i+q)^s H_{n,q,s}) - \log c \right| \right. \\
\left. - \frac{1}{c} \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \left| p_i (i+q)^s H_{n,q,s} - c \right| \right|. \qquad (5.77)
\end{aligned}
$$

Proof. Using (5.41) for $\phi(x) = -\log x$ and then putting $q_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,...,n$, we obtain (5.112). □

Remark 5.5 If we set $c = P_n$ in (5.112) then we obtain the inequality (14) as obtained in [4].

Another application of Theorem 5.23 for Zipf-Mandelbrot entropies corresponding to different parameters has been given below.

Theorem 5.42 ([5]) *Let $n \in \{1,2,3,...\}$, $t_1,t_2 \geq 0$, $s_1,s_2 > 0$, then for any $c \in \mathbb{R}_+$, we have*

$$\sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}} - Z(H,t_2,s_2) + \log c + (1-c)\frac{1}{c}$$
$$\geq \left| \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \log \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - \log c \right| \right.$$
$$\left. - \frac{1}{c} \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - c \right| \right|. \tag{5.78}$$

Proof. Using (5.41) for $\phi(x) = -\log x$, we obtain

$$\frac{-1}{Q_n} \sum_{i=1}^{n} q_i \log \frac{p_i}{q_i} + \log c + \left(\frac{P_n}{Q_n} - c \right) \frac{1}{c}$$
$$\geq \left| \frac{1}{Q_n} \sum_{i=1}^{n} q_i \left| \log \frac{p_i}{q_i} - \log c \right| - \frac{1}{cQ_n} \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|. \tag{5.79}$$

Now submitting p_i by $\frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and q_i by $\frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,...,n$, in (5.79) we deduce (5.112). □

Remark 5.6 If we set $c = 1$ in (5.78) then we obtain the inequality (16) as obtained in [4].

In the rest of results we have obtained related results by using another convex functions.

Theorem 5.43 ([5]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $s, q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} q_i = Q_n$, then for any $c \in \mathbb{R}_+$, we have*

$$Q_n c \log c + (1 - cQ_n)(1 + \log c) \leq -\sum_{i=1}^{n} \frac{\log q_i}{(i+q)^s H_{n,q,s}} - Z(H,q,s). \tag{5.80}$$

Proof. Using $\phi(x) = x \log x$ in the first inequality of (5.34) we obtain

$$Q_n c \log c + (P_n - cQ_n)(1 + \log c) \leq \sum_{i=1}^{n} p_i \log \left(\frac{p_i}{q_i} \right). \tag{5.81}$$

Let $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,\ldots,n$, then $P_n = 1$ and proceeding as in the proof of Theorem 5.50, we have

$$\sum_{i=1}^{n} p_i \log p_i = -Z(H,q,s).$$

Therefore (5.85) implies that

$$Q_n c \log c + (1 - cQ_n)(1 + \log c) \le -\sum_{i=1}^{n} \frac{\log q_i}{(i+q)^s H_{n,q,s}} - Z(H,q,s). \tag{5.82}$$

□

Corollary 5.23 ([5]) *Let $n \in \{1,2,3,\ldots\}$, $q \ge 0$, $s, q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} q_i = Q_n$, then for any $c \in \mathbb{R}_+$, we have*

$$\log \frac{1}{Q_n} \le -\sum_{i=1}^{n} \frac{\log q_i}{(i+q)^s H_{n,q,s}} - Z(H,q,s). \tag{5.83}$$

Proof. By substituting $c = \frac{1}{Q_n}$ in (5.57), we get (5.83). □

Theorem 5.44 ([5]) *Let $n \in \{1,2,3,\ldots\}$, $t_1, t_2 \ge 0$, $s_1, s_2 > 0$, then for any $c \in \mathbb{R}_+$, we have*

$$c \log c + (1-c)(1 + \log c) \le \sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - Z(H,t_1,s_1) \tag{5.84}$$

Proof. Using $\phi(x) = x \log x$ in the first inequality of (5.34) we obtain

$$Q_n c \log c + (P_n - cQ_n)(1 + \log c) \le \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right). \tag{5.85}$$

Using (5.85) for $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,\ldots,n$, we obtain

$$\sum_{i=1}^{n} p_i \log p_i = -Z(H,t_1,s_1), \sum_{i=1}^{n} p_i \log q_i = -\sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}}.$$

Therefore (5.85) implies that

$$c \log c + (1-c)(1 + \log c) \le \sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - Z(H,t_1,s_1).$$

This completes the proof. □

Corollary 5.24 ([5]) *Let $n \in \{1,2,3,\ldots\}$, $t_1, t_2 \ge 0$, $s_1, s_2 > 0$, then we have*

$$Z(H,t_1,s_1) \le \sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \tag{5.86}$$

Proof. By taking $c = 1$ in (5.84) we get (5.86). □

Theorem 5.45 ([5]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $s, p_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} p_i = P_n$, then for any $d \in \mathbb{R}_+$, we have*

$$\begin{aligned} &d\log d + P_n - d\Big(1 + \sum_{i=1}^{n} \frac{\log p_i}{(i+q)^s H_{n,q,s}} + Z(H,q,s)\Big) \\ &\geq \Bigg| \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \Big| d\log d - p_i (i+q)^s H_{n,q,s} \log p_i (i+q)^s H_{n,q,s} \Big| \\ &\quad - \sum_{i=1}^{n} (i+q)^s H_{n,q,s} \Big| (1 + \log p_i (i+q)^s H_{n,q,s})(d - p_i (i+q)^s H_{n,q,s}) \Big| \Bigg|. \quad (5.87) \end{aligned}$$

Proof. Using $\phi(x) = x\log x$ in the inequality (5.38), then we obtain

$$\begin{aligned} &d\log d - \frac{1}{Q_n} \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} - \frac{1}{Q_n} \sum_{i=1}^{n} q_i \left(1 + \log \frac{p_i}{q_i}\right)\left(d - \frac{p_i}{q_i}\right) \\ &\geq \left| \frac{1}{Q_n} \sum_{i=1}^{n} q_i \Big| d\log d - \frac{p_i}{q_i} \log \frac{p_i}{q_i} \Big| - \frac{1}{Q_n} \sum_{i=1}^{n} q_i \Big| (1 + \log \frac{p_i}{q_i})(d - \frac{p_i}{q_i}) \Big| \right|. \quad (5.88) \end{aligned}$$

which is equivalent to

$$\begin{aligned} &d\log d + \frac{P_n}{Q_n} - d\Big(1 + \sum_{i=1}^{n} q_i \log \frac{p_i}{q_i}\Big) \\ &\geq \left| \frac{1}{Q_n} \sum_{i=1}^{n} q_i \Big| d\log d - \frac{p_i}{q_i} \log \frac{p_i}{q_i} \Big| - \frac{1}{Q_n} \sum_{i=1}^{n} q_i \Big| (1 + \log \frac{p_i}{q_i})(d - \frac{p_i}{q_i}) \Big| \right|. \quad (5.89) \end{aligned}$$

Putting $q_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,...,n$, in (5.89), we get (5.87). □

Theorem 5.46 ([5]) *Let $n \in \{1,2,3,...\}$, $t_1, t_2 \geq 0$, $s_1, s_2 > 0$, then for any $d \in \mathbb{R}_+$, we have*

$$\begin{aligned} &d\log d + 1 - d\Big(1 - \sum_{i=1}^{n} \frac{\log((i+t_1)^{s_1} H_{n,t_1,s_1})}{(i+t_2)^{s_2} H_{n,t_2,s_2}} + Z(H,t_2,s_2)\Big) \\ &\geq \Bigg| \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \Big| d\log d - \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \Big| \\ &\quad - \sum_{i=1}^{n} \frac{1}{(i+t_1)^{s_1} H_{n,t_2,s_2}} \Big| (1 + \log \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}})(d - \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}}) \Big| \Bigg|. \quad (5.90) \end{aligned}$$

Proof. Using (5.85) for $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,...,n$, we obtain (5.90). □

Theorem 5.47 ([5]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $s, q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} q_i = Q_n$, then for any $c \in \mathbb{R}_+$, we have*

$$-Z(H,q,s) - \sum_{i=1}^{n} \frac{\log q_i}{(i+q)^s H_{n,q,s}} - Q_n c \log c - (1 - cQ_n)(1 + \log c)$$
$$\geq \left| \sum_{i=1}^{n} q_i \left| \frac{\log(q_i (i+q)^s H_{n,q,s})}{q_i (i+q)^s H_{n,q,s}} + c \log c \right| - |1 + \log c| \sum_{i=1}^{n} q_i \left| \frac{1}{q_i (i+q)^s H_{n,q,s}} - c \right| \right|. \tag{5.91}$$

Proof. Using inequality (5.41) for $\phi(x) = x \log x$ we obtain

$$\frac{1}{Q_n} \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} - c \log c - \left(\frac{P_n}{Q_n} - c \right)(1 + \log c)$$
$$\geq \left| \frac{1}{Q_n} \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} \log \frac{p_i}{q_i} - c \log c \right| - \frac{|1 + \log c|}{Q_n} \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|. \tag{5.92}$$

Now putting $p_i = \frac{1}{(i+q)^s H_{n,q,s}}$, $i = 1,2,...,n$, in (5.92), we get (5.87). □

Remark 5.7 If we set $c = \frac{1}{Q_n}$ in (5.87) we deduce the inequality (17) as obtained in [4].

Theorem 5.48 ([5]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $s, q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} q_i = Q_n$, then for any $c \in \mathbb{R}_+$, we have*

$$\sum_{i=1}^{n} \frac{\log((i+t_2)^{s_2} H_{n,t_2,s_2})}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - Z(H,t_1,s_1) - c \log c - (1-c)(1 + \log c)$$
$$\geq \left| \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - c \log c \right| - |1 + \log c| \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \left| \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} - c \right| \right|. \tag{5.93}$$

Proof. Using (5.92) for $p_i = \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $q_i = \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, $i = 1,2,...,n$, we get (5.93). □

Remark 5.8 If we set $c = 1$ in (5.93) we deduce the inequality (19) as obtained in [4].

Now we give applications of the above some results in Linguistics.

Gelbukh and Sidorov in [14] observed the difference between the coefficients s_1 and s_2 in Zipf's law for the English and Russian languages. They processed 39 literature texts for each language, chosen randomly from different genres, with the requirement that the size be greater than 10,000 running words each. They calculated coefficients for each of the mentioned texts and as the result they obtained the average $s_1 = 0.973863$ for the English language and $s_2 = 0.892869$ for the Russian language.

In the following result we give application of inequalities (5.61) for the English language.

Application 5.1 [[5]] Let $n \in \{1,2,3,...\}$, $p_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^n p_i = P_n$, then for any $c,d \in \mathbb{R}_+$, we have

$$\log c + (P_n - c)\frac{1}{c} \geq \sum_{i=1}^n \frac{\log p_i}{i^{0.973863} H_{n,0,0.973863}} + Z(H,0,0.973863)$$
$$\geq \log d + 1 - d\sum_{i=1}^n \frac{1}{i^{1.947726} p_i H_{n,0,0.973863}^2}. \tag{5.94}$$

Similarly we can give application for Russian language.

Now we give application of the result related two parameters: $s_1 = 0.973863$ for the English language and $s_2 = 0.892869$ for the Russian language, which is in fact application of the inequalities in (5.65).

Application 5.2 [[5]] Let $n \in \{1,2,3,...\}$, then for any $c,d \in \mathbb{R}_+$, we have

$$\log c + (1 - c)\frac{1}{c} \geq Z(H,0,0.892869) - \sum_{i=1}^n \frac{\log(i^{0.973863} H_{n,0,0.973863})}{i^{0.892869} H_{n,0,0.892869}}$$
$$\geq \log d + 1 - d\sum_{i=1}^n \frac{i^{0.973863} H_{n,0,0.973863}}{i^{0.892869} H_{n,0,0.892869})^2}. \tag{5.95}$$

Remark 5.9 By the similar way we can give applications of other related results.

5.7 Results for hybrid Zipf-Mandelbrot entropy

We start to give first general inequalities for Hybrid Zipf-Mandelbrot entropy which contain two arbitrary positive real numbers.

Theorem 5.49 ([6]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $w,s,p_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^n p_i = P_n$, then for any $c,d \in \mathbb{R}_+$, we have*

$$\log c + (P_n - c)\frac{1}{c} \geq \sum_{i=1}^n \frac{w^i \log p_i}{(i+q)^s \Phi^*(s,q,w)} + H(\Phi^*,q,s)$$
$$\geq \log d + 1 - d\sum_{i=1}^n \frac{w^{2i}}{p_i (i+q)^{2s} \Phi^{*2}(s,q,w)}, \tag{5.96}$$

Proof. Taking discrete measure, $p \to p_i$, $q \to q_i$ and then taking $\phi(x) = -\log x$ in (5.43) we obtain

$$-Q_n \log c - (P_n - cQ_n)\frac{1}{c} \le \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right) \le -Q_n \log d - Q_n + d\sum_{i=1}^{n} \frac{q_i^2}{p_i}. \tag{5.97}$$

Let $q_i = \frac{w^i}{(i+q)^s \Phi^*(s,q,w)}$, $i = 1,2,\ldots,n$, then $Q_n = \sum_{i=1}^{n} q_i = 1$ and

$$\begin{aligned}
\sum_{i=1}^{n} q_i \log q_i &= \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \log \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \\
&= \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \left(\log \frac{w^i}{(i+q)^s} + \log \frac{1}{\Phi^*(s,q,w)}\right) \\
&= \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \log \frac{w^i}{(i+q)^s} - \frac{\log \Phi^*(s,q,w)}{\Phi^*(s,q,w)} \sum_{i=1}^{n} \frac{w^i}{(i+q)^s} = -H(\Phi^*,q,s)
\end{aligned}$$

and

$$\sum_{i=1}^{n} q_i \log p_i = \sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s \Phi^*(s,q,w)}$$

Therefore (5.97) implies that

$$\begin{aligned}
-\log c - (P_n - c)\frac{1}{c} &\le -\sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s \Phi^*(s,q,w)} - H(\Phi^*,q,s) \\
&\le -\log d - 1 + d\sum_{i=1}^{n} \frac{w^{2i}}{p_i (i+q)^{2s} \Phi^{*2}(s,q,w)},
\end{aligned} \tag{5.98}$$

which is equivalent to (5.96). □

The following corollaries are consequences of the above theorem.

Corollary 5.25 ([6]) *Let $n \in \{1,2,3,\ldots\}$, $q \ge 0$, $w,s,p_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} p_i = P_n$, then we have*

$$\begin{aligned}
0 &\ge \sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s \Phi^*(s,q,w)} + H(\Phi^*,q,s) - \log P_n \\
&\ge 1 - P_n \sum_{i=1}^{n} \frac{w^{2i}}{p_i (i+q)^{2s} \Phi^{*2}(s,q,w)}.
\end{aligned} \tag{5.99}$$

Proof. Taking $c = d = P_n$ in (5.117), we get (5.99). □

The following consequence of Theorem 5.50 is in fact the application of Slater's inequality.

Corollary 5.26 ([6]) *Let $n \in \{1,2,3,\ldots\}$, $q \ge 0$, $w,s,p_i > 0$, $i = 1,2,..,n$, then*

$$\sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s \Phi^*(s,q,w)} + H(\Phi^*,q,s) \ge \log \frac{1}{\sum_{i=1}^{n} \frac{w^{2i}}{p_i (i+q)^{2s} \Phi^{*2}(s,q,w)}}. \tag{5.100}$$

Proof. By setting $d = \frac{1}{\sum_{i=1}^{n} \frac{w^{2i}}{p_i (i+q)^{2s} \Phi^{*2}(s,q,w)}}$ in the right inequality of (5.117), we get (5.100). □

In the following theorem we give general inequalities for hybrid Zipf-Mandelbrot entropies corresponding to different parameters.

Theorem 5.50 ([6]) *Let $n \in \{1,2,3,...\}$, $t_1, t_2 \geq 0$, $w_1, w_2, s_1, s_2 > 0$, then for any $c, d \in \mathbb{R}_+$, we have*

$$\begin{aligned} \log c + (1-c)\frac{1}{c} &\geq H(\Phi^*, t_2, s_2) + \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} \log \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} \\ &\geq \log d + 1 - d\sum_{i=1}^{n} \frac{w_2^{2i}(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}{w_1^i((i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2))^2}. \end{aligned} \tag{5.101}$$

Proof. Let $p_i = \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}$ and $q_i = \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}$, $i = 1,2,...,n$, then

$$\sum_{i=1}^{n} q_i \log q_i = \sum_{i=1}^{n} \frac{1}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} \log \frac{1}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} = -H(\Phi^*, t_2, s_2),$$

$$\sum_{i=1}^{n} q_i \log p_i = \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} \log \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}$$

Also $Q_n = \sum_{i=1}^{n} q_i = \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} = 1$ and $P_n = \sum_{i=1}^{n} \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} = 1$

Therefore using (5.97) for $p_i = \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}$ and $q_i = \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}$, $i = 1,2,..,n$, we obtain (5.101). □

The following corollaries are the consequence of Theorem 5.50.

Corollary 5.27 ([6]) *Let $n \in \{1,2,3,...\}$, $t_1, t_2 \geq 0$, $w_1, w_2, s_1, s_2 > 0$, then*

$$\begin{aligned} 0 &\geq H(\Phi^*, t_2, s_2) + \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} \log \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} \\ &\geq 1 - \sum_{i=1}^{n} \frac{w_2^{2i}(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}{w_1^i((i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2))^2}. \end{aligned} \tag{5.102}$$

Proof. Take $c = d = 1$ in (5.101), we get (5.102). □

The following corollary is the application of Theorem 5.50 which is in fact the application of Slater's inequality for hybrid Zipf-Mandelbrot entropies corresponding to different parameters.

Corollary 5.28 ([6]) *Let* $n \in \{1,2,3,...\}$, $t_1,t_2 \geq 0$, $w_1,w_2,s_1,s_2 > 0$, *then*

$$\sum_{i=1}^{n} \frac{w_2^i \log\left(\frac{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}{w_1^i}\right)}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} - H(\Phi^*,t_2,s_2)$$
$$\leq \log\left(\sum_{i=1}^{n} \frac{w_2^{2i}(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}{w_1^i((i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2))^2}\right). \tag{5.103}$$

Proof. Take $d = \frac{1}{\sum_{i=1}^{n} \frac{w_2^{2i}(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}{w_1^i((i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2))^2}}$ in the right inequality of (5.101), we get (5.103).

□

The following result for hybrid Zipf-Mandelbrot entropy has been obtained by using the right inequality in (5.46).

Theorem 5.51 ([6]) *Let* $n \in \{1,2,3,...\}$, $q \geq 0$, $w,s,p_i > 0$, $i = 1,2,..,n$, *then for any* $d \in \mathbb{R}_+$ *we have*

$$\sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s\Phi^*(s,q,w)} + H(\Phi^*,q,s) \geq \log\left(\frac{1}{\sum_{i=1}^{n} \frac{w^{2i}}{p_i(i+q)^{2s}\Phi^{*2}(s,q,w)}}\right)$$
$$\geq \log d + 1 - d\sum_{i=1}^{n} \frac{w^{2i}}{p_i(i+q)^{2s}\Phi^{*2}(s,q,w)}. \tag{5.104}$$

Proof. Taking discrete measure, $p \longrightarrow p_i$, $q \longrightarrow q_i$ and then taking $\phi(x) = -\log x$ in (5.43) we obtain

$$\frac{-1}{Q_n}\sum_{i=1}^{n} q_i \log\frac{p_i}{q_i} \leq -\log\frac{Q_n}{\sum_{i=1}^{n}\frac{q_i^2}{p_i}} \leq -\log d - \frac{1}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i}\left(\frac{p_i}{q_i} - d\right). \tag{5.105}$$

Now putting $q_i = \frac{w^i}{(i+q)^s\Phi^*(s,q,w)}$, $i = 1,2,...,n$, in (5.105) we deduce

$$-\sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s\Phi^*(s,q,w)} - H(\Phi^*,q,s) \leq -\log\left(\frac{1}{\sum_{i=1}^{n} \frac{w^{2i}}{p_i(i+q)^{2s}\Phi^{*2}(s,q,w)}}\right)$$
$$\leq -\log d - 1 + d\sum_{i=1}^{n} \frac{w^{2i}}{p_i(i+q)^{2s}\Phi^{*2}(s,q,w)},$$

which is equivalent to (5.104). □

In the following theorem we have presented result for two hybrid Zipf-Mandelbrot entropies corresponding to different parameters.

Theorem 5.52 ([6]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $w_1,w_2,s_1,s_2 > 0$, then for any $d \in \mathbb{R}_+$, we have*

$$\sum_{i=1}^{n} \frac{w_2^i \log\left(\frac{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}{w_1^i}\right)}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} - H(\Phi^*,t_2,s_2) \leq \log \sum_{i=1}^{n} \frac{((i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2))^2}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$$
$$\leq -\log d - 1 + \frac{d w_2^{2i}(i+t_1)^{s_1} H_{n,t_1,s_1}}{w_1^i((i+t_2)^{s_2} H_{n,t_2,s_2})^2}. \tag{5.106}$$

Proof. As in the proof of Theorem 5.51, putting $p_i = \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}$ and $q_i = \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}$, $i = 1,2,\dots,n$, in (5.105), we get (5.106). □

The following application of the inequality (5.47) for hybrid Zipf-Mandelbrot entropy holds.

Theorem 5.53 ([6]) *Let $n \in \{1,2,3,\dots\}$, $q \geq 0$, $w,s,p_i > 0$, $i = 1,2,..,n$, then for any $d \in \mathbb{R}_+$, we have*

$$-\log d + \sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s \Phi^*(s,q,w)} + H(\Phi^*,t,s) - 1 + d \sum_{i=1}^{n} \frac{w^{2i}}{p_i (i+q)^{2s} \Phi^{*2}(s,q,w)}$$
$$\geq \left| \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \left| \log d - \log \frac{p_i (i+q)^s \Phi^*(s,q,w)}{w^i} \right| \right.$$
$$\left. - \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \left| \frac{d w^i}{p_i (i+q)^s \Phi^*(s,q,w)} - 1 \right| \right|. \tag{5.107}$$

Proof. Taking discrete measure, $p \to p_i$, $q \to q_i$ and then taking $\phi(x) = -\log x$ in (5.47) we obtain

$$-\log d + \sum_{i=1}^{n} q_i \log \frac{p_i}{q_i} + \frac{1}{Q_n} \sum_{i=1}^{n} \frac{q_i^2}{p_i} \left(d - \frac{p_i}{q_i}\right)$$
$$\geq \left| \frac{1}{Q_n} \sum_{i=1}^{n} q_i \left| \log d - \log \frac{p_i}{q_i} \right| - \frac{1}{Q_n} \sum_{i=1}^{n} q_i \left| \frac{q_i}{p_i} (d - \frac{p_i}{q_i}) \right| \right|. \tag{5.108}$$

Now putting $q_i = \frac{w^i}{(i+q)^s \Phi^*(s,q,w)}$, $i = 1,2,\dots,n$, in (5.108) we obtain (5.107). □

The following consequence of the above theorem is in fact the refinement of inequality (5.100).

Corollary 5.29 ([6]) *Let $n \in \{1,2,3,\dots\}$, $q \geq 0$, $w,s,p_i > 0$, $i = 1,2,..,n$ and $\tilde{x} := \frac{1}{\sum_{i=1}^{n} \frac{w^{2i}}{p_i (i+q)^{2s} \Phi^{*2}(s,q,w)}}$. Then*

$$-\log \tilde{x} + \sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s \Phi^*(s,q,w)} + H(\Phi^*,t,s)$$

$$\geq \left| \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \left| \log \tilde{x} - \log \frac{p_i (i+q)^s \Phi^*(s,q,w)}{w^i} \right| \right.$$
$$\left. - \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \left| \frac{\tilde{x} w^i}{p_i (i+q)^s \Phi^*(s,q,w)} - 1 \right| \right|. \tag{5.109}$$

Proof. By setting $d = \tilde{x}$ in (5.107) we get (5.109). □

In the following theorem we have given refinement of the right inequality in (5.101).

Theorem 5.54 ([6]) *Let* $n \in \{1,2,3,...\}$, $w_1, w_2, t_1, t_2 \geq 0$, $s_1, s_2 > 0$, *then for any* $d \in \mathbb{R}_+$, *we have*

$$H(\Phi^*, t_2, s_2) + \sum_{i=1}^{n} \frac{w_2^i \log \frac{w_1^i}{(i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)}}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)} - \log d - 1 + d \sum_{i=1}^{n} \frac{w_2^{2i} (i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)}{w_1^i ((i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2))^2}$$
$$\geq \left| \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)} \left| \log d - \log \frac{w_1^i (i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)} \right| \right.$$
$$\left. - \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)} \left| \frac{d w_1^i (i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)} - 1 \right| \right|. \tag{5.110}$$

Proof. Putting $p_i = \frac{w_1^i}{(i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)}$ and $q_i = \frac{w_2^i}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}$, $i = 1,2,...,n$, in (5.108), we obtain (5.110). □

The following consequence of the above theorem is in fact refinement of Slater's inequality for Zipf-Mandelbrot entropies corresponding to different parameters.

Corollary 5.30 ([6]) *Let* $n \in \{1,2,3,...\}$, $t_1, t_2 \geq 0$, $s_1, s_2 > 0$ *and* $\tilde{x} := \frac{1}{\sum_{i=1}^{n} \frac{w_2^i (i+t_1)^{s_1} H_{n,t_1,s_1}}{w_1^i ((i+t_2)\Phi^*(t_2,s_2,w_2))^2}}$,

then

$$-\log \tilde{x} + Z(H, t_2, s_2) + \sum_{i=1}^{n} \frac{w_2^i \log \frac{w_1^i}{(i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)}}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}$$
$$\geq \left| \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)} \left| \log \tilde{x} - \log \frac{w_1^i (i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)} \right| \right.$$
$$\left. - \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)} \left| \frac{\tilde{x} w_1^i (i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)} - 1 \right| \right|. \tag{5.111}$$

Proof. By setting $d = \frac{1}{\sum_{i=1}^{n} \frac{w_2^i (i+t_1)^{s_1} H_{n,t_1,s_1}}{w_1^i ((i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2))^2}}$ in (5.110), we obtain (5.111). □

The following theorem is the application of Theorem 5.28 for hybrid Zipf-Mandelbrot entropy.

Theorem 5.55 ([6]) *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $w,s,p_i > 0$, $i = 1,2,..,n$, then for any $c \in \mathbb{R}_+$, we have*

$$
\begin{aligned}
&-\sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s \Phi^*(s,q,w)} - H(\Phi^*,t,s) + \log c + (P_n - c)\frac{1}{c} \\
&\geq \left| \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \left| \log\left(\frac{p_i (i+q)^s \Phi^*(s,q,w)}{w^i}\right) - \log c \right| \right. \\
&\qquad \left. - \frac{1}{c} \sum_{i=1}^{n} \frac{w^i}{(i+q)^s \Phi^*(s,q,w)} \left| \frac{p_i (i+q)^s \Phi^*(s,q,w)}{w^i} - c \right| \right|. \qquad (5.112)
\end{aligned}
$$

Proof. Taking discrete measure, $p \to p_i$, $q \to q_i$ and then taking $\phi(x) = -\log x$ in (5.50) we get

$$
\begin{aligned}
&-\frac{1}{Q_n} \sum_{i=1}^{n} q_i \log\left(\frac{p_i}{q_i}\right) + \log c + \left(\frac{P_n}{Q_n} - c\right)\frac{1}{c} \\
&\geq \left| \frac{1}{Q_n} \sum_{i=1}^{n} q_i \left| \log\left(\frac{p_i}{q_i}\right) - \log c \right| - \frac{1}{cQ_n} \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|. \qquad (5.113)
\end{aligned}
$$

Now putting $q_i = \frac{w^i}{(i+q)^s \Phi^*(s,q,w)}$, $i = 1,2,\ldots,n$, in (5.113) we obtain (5.112). □

Another application of Theorem 5.28 for Zipf-Mandelbrot entropies corresponding to different parameters has been given below.

Theorem 5.56 ([6]) *Let $n \in \{1,2,3,\ldots\}$, $t_1,t_2 \geq 0$, $w_1,w_2,s_1,s_2 > 0$, then for any $c \in \mathbb{R}_+$, we have*

$$
\begin{aligned}
&\log c + (1-c)\frac{1}{c} - H(\Phi^*,t_2,s_2) + \sum_{i=1}^{n} \frac{w_2^i \log \frac{w_1^i}{(i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)}}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)} \\
&\geq \left| \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)} \left| \log \frac{w_1^i (i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)} - \log c \right| \right. \\
&\left. - \frac{1}{c} \sum_{i=1}^{n} \frac{w_2^i}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)} \left| \frac{w_1^i (i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)} - c \right| \right|. \qquad (5.114)
\end{aligned}
$$

Proof. Substitute $p_i = \frac{w_1^i}{(i+t_1)^{s_1} \Phi^*(t_1,s_1,w_1)}$ and $q_i = \frac{w_2^i}{(i+t_2)^{s_2} \Phi^*(t_2,s_2,w_2)}$, $i = 1,2,\ldots,n$, in (5.113) we deduce (5.114). □

In the rest of results we have obtained related results by using another convex functions.

Theorem 5.57 ([6]) *Let $n \in \{1,2,3,\ldots\}$, $q \geq 0$, $w,s,q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} q_i = Q_n$, then for any $c \in \mathbb{R}_+$, we have*

$$
Q_n c \log c + (1 - cQ_n)(1 + \log c) \leq -\sum_{i=1}^{n} \frac{w^i \log q_i}{(i+q)^s \Phi^*(s,q,w)} - H(\Phi^*,t_2,s_2). \qquad (5.115)
$$

Proof. Taking discrete measure, $p \to p_i$, $q \to q_i$ and then using $\phi(x) = x\log x$ in the first inequality of (5.43) we obtain

$$Q_n c\log c + (P_n - cQ_n)(1+\log c) \le \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right). \tag{5.116}$$

Let $p_i = \frac{w^i}{(i+q)^s \Phi^*(s,q,w)}$, $i = 1,2,\ldots,n$, then $P_n = 1$ and proceeding as in the proof of Theorem 5.50, we have

$$\sum_{i=1}^{n} p_i \log p_i = -H(\Phi^*, t_2, s_2).$$

Therefore (5.116) implies that

$$Q_n c\log c + (1 - cQ_n)(1+\log c) \le -\sum_{i=1}^{n} \frac{w^i \log q_i}{(i+q)^s \Phi^*(s,q,w)} - H(\Phi^*, t_2, s_2). \tag{5.117}$$

□

Corollary 5.31 ([6]) *Let $n \in \{1,2,3,\ldots\}$, $q \ge 0$, $w, s, q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} q_i = Q_n$, then we have*

$$\log\frac{1}{Q_n} \le -\sum_{i=1}^{n} \frac{w^i \log q_i}{(i+q)^s \Phi^*(s,q,w)} - H(\Phi^*, t_2, s_2). \tag{5.118}$$

Proof. By setting $c = \frac{1}{Q_n}$ in (5.115) we obtain (5.118). □

Theorem 5.58 ([6]) *Let $n \in \{1,2,3,\ldots\}$, $t_1, t_2 \ge 0$, $w_1, w_2, s_1, s_2 > 0$, then for any $c \in \mathbb{R}_+$, we have*

$$c\log c + (1-c)(1+\log c) \le \sum_{i=1}^{n} \frac{w_1^i \log\left(\frac{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i}\right)}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} - H(\Phi^*, t_2, s_2). \tag{5.119}$$

Proof. If $p_i = \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}$ and $q_i = \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}$, $i = 1,2,\ldots,n$, then

$$\sum_{i=1}^{n} p_i \log p_i = -H(\Phi^*, t_2, s_2), \sum_{i=1}^{n} p_i \log q_i = \sum_{i=1}^{n} \frac{w_1^i \log\left(\frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}\right)}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}.$$

Therefore (5.116) implies that

$$c\log c + (1-c)(1+\log c) \le \sum_{i=1}^{n} \frac{w_1^i \log\left(\frac{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i}\right)}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} - H(\Phi^*, t_2, s_2).$$

This completes the proof. □

Corollary 5.32 ([6]) *Let $n \in \{1,2,3,\dots\}$, $t_1,t_2 \geq 0$, $w_1,w_2,s_1,s_2 > 0$, then we have*

$$H(\Phi^*,t_2,s_2) \leq \sum_{i=1}^{n} \frac{w_1^i \log(\frac{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i})}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}. \tag{5.120}$$

Proof. By taking $c = 1$ in (5.119), we get (5.120). □

Theorem 5.59 ([6]) *Let $n \in \{1,2,3,\dots\}$, $q \geq 0$, $w,s,p_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} p_i = P_n$, then for any $d \in \mathbb{R}_+$, we have*

$$\begin{aligned}
&d\log d + P_n - d(1+\sum_{i=1}^{n} \frac{w^i \log p_i}{(i+q)^s\Phi^*(s,q,w)} + H(\Phi^*,t,s)) \\
&\geq \Bigg|\sum_{i=1}^{n} \frac{w^i}{(i+q)^s\Phi^*(s,q,w)}\Bigg|d\log d - \frac{p_i(i+q)^s\Phi^*(s,q,w)}{w^i}\log\left(\frac{p_i(i+q)^s\Phi^*(s,q,w)}{w^i}\right)\Bigg| \\
&-\sum_{i=1}^{n}(i+q)^s H_{n,q,s}\Bigg|\left(1+\log\left(\frac{p_i(i+q)^s\Phi^*(s,q,w)}{w^i}\right)\right)\left(d - \frac{p_i(i+q)^s\Phi^*(s,q,w)}{w^i}\right)\Bigg|\Bigg|.
\end{aligned} \tag{5.121}$$

Proof. Taking discrete measure, $p \to p_i$, $q \to q_i$ and then using $\phi(x) = x\log x$ in the inequality (5.47), then we obtain

$$\begin{aligned}
&d\log d - \frac{1}{Q_n}\sum_{i=1}^{n} p_i \log\frac{p_i}{q_i} - \frac{1}{Q_n}\sum_{i=1}^{n} q_i\left(1+\log\frac{p_i}{q_i}\right)\left(d - \frac{p_i}{q_i}\right) \\
&\geq \left|\frac{1}{Q_n}\sum_{i=1}^{n} q_i\left|d\log d - \frac{p_i}{q_i}\log\frac{p_i}{q_i}\right| - \frac{1}{Q_n}\sum_{i=1}^{n} q_i\left|(1+\log\frac{p_i}{q_i})(d - \frac{p_i}{q_i})\right|\right|.
\end{aligned} \tag{5.122}$$

which is equivalent to

$$\begin{aligned}
&d\log d + \frac{P_n}{Q_n} - d(1+\sum_{i=1}^{n} q_i \log\frac{p_i}{q_i}) \\
&\geq \left|\frac{1}{Q_n}\sum_{i=1}^{n} q_i\left|d\log d - \frac{p_i}{q_i}\log\frac{p_i}{q_i}\right| - \frac{1}{Q_n}\sum_{i=1}^{n} q_i\left|(1+\log\frac{p_i}{q_i})(d - \frac{p_i}{q_i})\right|\right|.
\end{aligned} \tag{5.123}$$

Putting $q_i = \frac{w^i}{(i+q)^s\Phi^*(s,q,w)}$, $i = 1,2,\dots,n$, in (5.123), we get (5.121). □

Theorem 5.60 ([6]) *Let $n \in \{1,2,3,...\}$, $t_1, t_2 \geq 0$, $w_1, w_2, s_1, s_2 > 0$, then for any $d \in \mathbb{R}_+$, we have*

$$d\log d + 1 - d(1 - \sum_{i=1}^{n} \frac{w_2^i \log\left(\frac{w_1^i (i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}\right)}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} + H(\Phi^*,t,s)))$$
$$\geq \left| \sum_{i=1}^{n} \frac{w_2^i \left| d\log d - \frac{w_1^i (i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} \log \frac{w_1^i (i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} \right|}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} \right.$$
$$\left. - \sum_{i=1}^{n} \frac{w_1^i \left| \left(1 + \log \frac{w_1^i (i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}\right)\left(d - \frac{w_1^i (i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}\right) \right|}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} \right|. \tag{5.124}$$

Proof. Using (5.123) for $p_i = \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}$ and $q_i = \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}$, $i = 1,2,...,n$, we obtain (5.124). □

Theorem 5.61 ([6]) *Let $n \in \{1,2,3,...\}$, $q \geq 0$, $w, s, q_i > 0$, $i = 1,2,..,n$ with $\sum_{i=1}^{n} q_i = Q_n$, then for any $c \in \mathbb{R}_+$, we have*

$$-H(\Phi^*,t,s) - \sum_{i=1}^{n} \frac{w^i \log q_i}{(i+q)^s\Phi^*(s,q,w)} - Q_n c\log c - (1 - cQ_n)(1 + \log c)$$
$$\geq \left| \sum_{i=1}^{n} q_i \left| \frac{w^i \log\left(\frac{w^i}{q_i (i+q)^s\Phi^*(s,q,w)}\right)}{q_i (i+q)^s\Phi^*(s,q,w)} - c\log c \right| \right.$$
$$\left. - |1 + \log c| \sum_{i=1}^{n} q_i \left| \frac{w^i}{q_i (i+q)^s\Phi^*(s,q,w)} - c \right| \right|. \tag{5.125}$$

Proof. Taking discrete measure, $p \to p_i$, $q \to q_i$ and then using inequality (5.50) for $\phi(x) = x\log x$, we obtain

$$\frac{1}{Q_n}\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} - c\log c - \left(\frac{P_n}{Q_n} - c\right)(1 + \log c)$$
$$\geq \left| \frac{1}{Q_n}\sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i}\log\frac{p_i}{q_i} - c\log c \right| - \frac{|1+\log c|}{Q_n}\sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - c \right| \right|. \tag{5.126}$$

Now putting $p_i = \frac{w^i}{(i+q)^s\Phi^*(s,q,w)}$, $i = 1,2,...,n$, in (5.126), we get (5.125). □

Theorem 5.62 ([6]) *Let* $n \in \{1,2,3,\dots\}$, $q \geq 0$, $w_1, w_2, s, q_i > 0$, $i = 1,2,..,n$ *with* $\sum_{i=1}^{n} q_i = Q_n$, *then for any* $c \in \mathbb{R}_+$, *we have*

$$\sum_{i=1}^{n} \frac{w_1^i \log\left(\frac{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i}\right)}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} - H(\Phi^*,t_1,s_1) - c\log c - (1-c)(1+\log c)$$
$$\geq \left| \sum_{i=1}^{n} \frac{w_2^i \left| \frac{w_1^i (i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} \log\left(\frac{w_1^i (i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}\right) - c\log c \right|}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} \right.$$
$$\left. - |1+\log c| \sum_{i=1}^{n} \frac{w_2^i \left| \frac{w_1^i (i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}{w_2^i (i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)} - c \right|}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)} \right|. \tag{5.127}$$

Proof. Using (5.126) for $p_i = \frac{w_1^i}{(i+t_1)^{s_1}\Phi^*(t_1,s_1,w_1)}$ and $q_i = \frac{w_2^i}{(i+t_2)^{s_2}\Phi^*(t_2,s_2,w_2)}$, $i = 1,2,\dots,n$, we get (5.127). □

Bibliography

[1] M. Adil Khan, M. Anwar, J. Jakšetić and J. Pečarić, *On some improvements of the Jensen inequality with some applications*, J. Inequal. Appl. **2009**, (2009), Article ID 323615, 15 pages.

[2] M. Adil Khan, S. Khalid and J. Pečarić,*Refinements of some majorization type inequalities*, J. Math. Inequal. **7**, 1 (2013), 73–92.

[3] M. Adil Khan, S. Khalid and J. Pečarić, *Improvement of Jensen's inequality in terms of the Gateaux derivatives for convex functions in linear spaces with applications*, Kyungpook Math. J. **52**, (2012), 495–511.

[4] M. Adil Khan, Đ. Pečarić and J. Pečarić, *Bounds for Shannon and Zipf-mandelbrot entropies*, Math. Meth. Appl. Sci., 2017;17. https://doi.org/10.1002/mma.4531.

[5] M. Adil Khan, Đ. Pečarić and J. Pečarić, On Zipf-Mandelbrot entropy, Journal of Computational and Applied Mathematics , **346** (2019) 192–204.

[6] M. Adil Khan, Đ. Pečarić and J. Pečarić, *Bounds for Shannon and hybrid Zipf-Mandelbrot entropies*, submitted.

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

[8] I. Csiszár and J. Korner, *Information theory: coding theorem for dicsrete memoryless systems*, Academic Press, New York (1981).

[9] L. Brillouin, *Science and Information Theory*, Academic Press, New York (1956).

[10] J. C. Willis, *Age and Area*, (Cambridge University Press, Cambridge, 1922).

[11] A. Koplenig, *Using the parameters of the Zipf-Mandelbrot law to measure diachronic lexical, syntactical and stylistic changes a large-scale corpus analysis*, Corpus Linguistics and inguistics Theory. (2015) (https://doi.org/10.1515/cllt-2014-0049).

[12] V. Diodato, *Dictionary of Bibliometrics*, Haworth Press, New York (1994).

[13] R. A. Fisher, *Theory of statistical estimation*, Proe. Cambridge Philos. Soc. **22**, (1925), 700–725.

[14] A. Gelbukh and G. Sidorov, Zipf and Heaps laws, Coefficients Depend on Language, Proc. CICLing 2001, Conference on Intelligent Text Processing and Computational Linguistics, Maxico City.

[15] J. Jakšetić, Đ. Pečarić and J. Pečarić, Hybrid Zipf-mandelbrot law, submitted.

[16] S. Kullback and A. S. Leibler, *On information and sufficiency*, Ann. Math. Statist. **22**, (1951), 79–86.

[17] C. E. Shannon, *A mathematical theory of communications*, Bell. Syst. Tech J. **27**, (1958), 623–659.

[18] B. Mandelbrot, *Information Theory and Psycholinguistics: A Theory of Words Frequencies. In Reading in Mathematical Social Scence. (ed.) P. Lazafeld*, N. Henry. Cambridge MA, MIT Press: (1966).

[19] A. Renyi, *On measures of entropy and information*, In Proc. Fourth Berkeley Syrup. on Math. Statist. and Prob. University of California Press: Berkeley, CA U.S.A, **1**, (1961), 547–561. https://projecteuclid.org/euclid.bsmsp/1200512181.

[20] L. Egghe and R. Rousseau, *Introduction to Informetrics, Quantitative Methods in Library, Documentation and Information Science*, Elsevier Science Publishers. New York: (1990).

[21] M. Matić and J. Pečarić, *Some companion inequalities to Jensen's inequality*, Math. Inequal. and Appl. **3**, 3 (2000), 355–368.

[22] S. S. Dragomir and C. J. Goh, *A counter part of Jensen's discrrete inequality for differntiable convex mapping and applications in information theory*, Math. Comput. Modelling. **24**, 2 (1996), 1–11.

[23] S. S. Dragomir, *A converse inequality for the Csiszar Φ-divergence*, Tamsui oxford J. Math. Sci. **20**, 1 (2004), 35–53.

[24] J. Pečarić, *A companion inequality to Jensen Steffensen's inequality*, J. Approx. Theory. **44**, (1985a), 289–291.

[25] S. S. Dragomir, *On a converse of Jensen's inequality*, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. **12** , (2001), 48–51.

[26] M. A. Montemurro, *Beyond the Zipf-Mandelbrot law in quantitative linguistics*, (2001) URL: arXiv:cond-mat/0104066v2 (22.2.2017.)

[27] D. Manin, *Mandelbrot's Model for Zipf's Law: Can Mandelbrot's Model Explain Zipf's Law for Language*, Journal. Quantitative. Linguistics. **16**, 3 (2009), 274–285.

[28] Z. K. Silagadze, *Citations and the Zipf-Mandelbrot Law*, Complex Systems. **11**, (1997), 487–499

[29] D. Mouillot and A. Lepretre, *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 (2000), 279–295.

[30] S. Hussain, J. Pečarić, *An improvement of Jensen's inequality with some applications*, Asian-Eur. J. Math. **2**, 1 (2009), 85–94.

[31] M. Visser, Zipf's law, power laws and maximum entropy, New J. Phys **15**, 043021 (2013).

Chapter 6

Combinatorial improvements of Zipf-Mandelbrot Laws via interpolations

Khuram Ali Khan, Tasadduq Nia, Đilda Pečarić and Josip Pečarić

Abstract. Jensens inequality is important to obtain inequalities for divergence between probability distributions. By applying a refinement of Jensen inequality [17] and introducing a new functional based on f-divergence functional, we obtain some estimates for the new functionals, the f-divergence and Rényi divergence. Some inequalities for Rényi and Shannon estimates are constructed. Zipf-Mandelbrot law is used to illustrate the results and generalize the refinement of Jensens inequality and new inequalities of Rényi and Shannon entropies for m-convex function using Montgomery identity, Lidstone polynomial, Taylor polynomial and Hermite interpolating polynomial.

2010 Mathematics Subject Classification. Primary 26D07, 26D15, 26D20, 26D99.
Key words and phrases. m-convex function, Jensen's Inequality, f- and Rényi divergence, Montgomery identity, Entropy.

6.1 Introduction

The most commonly used words, the largest cities of countries income of billionare can be described in term of Zipf's law. The f-divergence which means that distance between two probability distribution by making an average value, which is weighted by a specified function. As f-divergence, there are other probabilities distributions like Csiszar f-divergence [19, 20], some special case of which are Kullback-Leibler-divergence use to find the appropriate distance between the probability distribution (see [23, 24]). The notion of distance is stronger than divergence because it give the properties of symmetry and triangle inequalities. Probability theory has application in many fields and the divergence between probability distribution have many application in these fields.

Many natural phenomena's like distribution of wealth and income in a society, distribution of face book likes, distribution of football goals follows power law distribution (Zipf's Law). Like above phenomena's, distribution of city sizes also follow Power Law distribution. Auerbach [5] first time gave the idea that the distribution of city size can be well approximated with the help of Pareto distribution (Power Law distribution). This idea was well refined by many researchers but Zipf [13] worked significantly in this field. The distribution of city sizes is investigated by many scholars of the urban economics, like Rosen and Resnick [30] , Black and Henderson [7], Ioannides and Overman [18], Soo [31], Anderson and Ge [4] and Bosker et al. [8]. Zipf's law states that: "The rank of cities with a certain number of inhabitants varies proportional to the city sizes with some negative exponent, say that is close to unit". In other words, Zipf's Law states that the product of city sizes and their ranks appear roughly constant. This indicates that the population of the second largest city is one half of the population of the largest city and the third largest city equal to the one third of the population of the largest city and the population of n-th city is $\frac{1}{n}$ of the largest city population. This rule is called rank, size rule and also named as Zipf's Law. Hence Zip's Law not only shows that the city size distribution follows the Pareto distribution, but also show that the estimated value of the shape parameter is equal to unity.

In [21] L. Horváth et al. introduced some new functionals based on the f-divergence functionals, and obtained some estimates for the new functionals. They obtained f-divergence and Rényi divergence by applying a cyclic refinement of Jensen's inequality. They also construct some new inequalities for Rényi and Shannon entropies and used Zipf-Madelbrot law to illustrate the results.

Higher order convex function was introduced by T. Popoviciu (see [28, p. 15]). The inequalities involving higher order convex functions are used by physicists in higher dimensions problems. S. I. Butt *et al*. in their work stated that many of the results are not true for higher order convex functions which are true for convex functions, which convince us to study the results involving higher order convexity (see [10]). In [28, p. 16], the following criteria is given to check the m-convexity of the function.
If $f^{(m)}$ exists, then f is m-convex if and only if $f^{(m)} \geq 0$.
In recent years many researchers have generalized the inequalities for m-convex func-

tions; like S. I. Butt et al. generalized the Popoviciu inequality for m-convex function using Taylor's formula, Lidstone polynomial, montgomery identity, Fink's identity, Abel-Gonstcharoff interpolation and Hermite interpolating polynomial (see [9, 10, 11, 12, 13]).

Since many years Jensen's inequality has of great interest. The researchers have given the refinement of Jensen's inequality by defining some new functions (see [16, 17]). Like many researchers L. Horváth and J. Pečarić in ([14, 17], see also [15, p. 26]), gave a refinement of Jensen's inequality for convex function. They defined some essential notions to prove the refinement given as follows:
Let X be a set, and:
$P(X) :=$ Power set of X,
$|X| :=$ Number of elements of X,
$\mathbb{N} :=$ Set of natural numbers with 0.
Consider $q \geq 1$ and $r \geq 2$ be fixed integers. Define the functions

$$F_{r,s} : \{1,\ldots,q\}^r \to \{1,\ldots,q\}^{r-1} \quad 1 \leq s \leq r,$$

$$F_r : \{1,\ldots,q\}^r \to P\left(\{1,\ldots,q\}^{r-1}\right),$$

and

$$T_r : P(\{1,\ldots,q\}^r) \to P\left(\{1,\ldots,q\}^{r-1}\right),$$

by

$$F_{r,s}(i_1,\ldots,i_r) := (i_1,i_2,\ldots,i_{s-1},i_{s+1},\ldots,i_r) \quad 1 \leq s \leq r,$$

$$F_r(i_1,\ldots,i_r) := \bigcup_{s=1}^{r} \{F_{r,s}(i_1,\ldots,i_r)\},$$

and

$$T_r(I) = \begin{cases} \phi, & I = \phi; \\ \bigcup\limits_{(i_1,\ldots,i_r)\in I} F_r(i_1,\ldots,i_r), & I \neq \phi. \end{cases}$$

Next let the function

$$\alpha_{r,i} : \{1,\ldots,q\}^r \to \mathbb{N} \quad 1 \leq i \leq q$$

defined by

$\alpha_{r,i}(i_1,\ldots,i_r)$ is the number of occurences of i in the sequence $(i_1,\ldots,i_r)$.

For each $I \in P(\{1,\ldots,q\}^r)$ let

$$\alpha_{I,i} := \sum_{(i_1,\ldots,i_r)\in I} \alpha_{r,i}(i_1,\ldots,i_r) \quad 1 \leq i \leq q.$$

(H_1) Let n,m be fixed positive integers such that $n \geq 1$, $m \geq 2$ and let I_m be a subset of $\{1,\ldots,n\}^m$ such that

$$\alpha_{I_m,i} \geq 1 \quad 1 \leq i \leq n.$$

Introduce the sets $I_l \subset \{1,\ldots,n\}^l (m-1 \geq l \geq 1)$ inductively by

$$I_{l-1} := T_l(I_l) \quad m \geq l \geq 2.$$

Obviously the sets $I_1 = \{1,\ldots,n\}$, by (H_1) and this insures that $\alpha_{I_1,i} = 1 (1 \leq i \leq n)$. From (H_1) we have $\alpha_{I_l,i} \geq 1 (m-1 \geq l \geq 1, 1 \leq i \leq n)$.
For $m \geq l \geq 2$, and for any $(j_1,\ldots,j_{l-1}) \in I_{l-1}$, let

$$\mathscr{H}_{I_l}(j_1,\ldots,j_{l-1}) := \{((i_1,\ldots,i_l),k) \times \{1,\ldots,l\} | F_{l,k}(i_1,\ldots,i_l) = (j_1,\ldots,j_{l-1})\}.$$

With the help of these sets they define the functions $\eta_{I_m,l} : I_l \to \mathbb{N}(m \geq l \geq 1)$ inductively by

$$\eta_{I_m,m}(i_1,\ldots,i_m) := 1 \quad (i_1,\ldots,i_m) \in I_m;$$
$$\eta_{I_m,l-1}(j_1,\ldots,j_{l-1}) := \sum_{((i_1,\ldots,i_l),k) \in \mathscr{H}_{I_l}(j_1,\ldots,j_{l-1})} \eta_{I_m,l}(i_1,\ldots,i_l).$$

They define some special expressions for $1 \leq l \leq m$, as follows

$$\mathscr{A}_{m,l} = \mathscr{A}_{m,l}(I_m, x_1,\ldots,x_n, p_1,\ldots,p_n; f) := \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l) \in I_l} \eta_{I_m,l}(i_1,\ldots,i_l)$$
$$\left(\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}} \right) f \left(\frac{\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}} x_{i_j}}{\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}}} \right)$$

and prove the following theorem.

Theorem 6.1 *Assume (H_1), and let $f : I \to \mathbb{R}$ be a convex function where $I \subset \mathbb{R}$ is an interval. If $x_1,\ldots,x_n \in I$ and $p_1,\ldots,p_n$ are positive real numbers such that $\sum_{i=1}^{n} p_i = 1$, then*

$$f\left(\sum_{s=1}^{n} p_s x_s\right) \leq \mathscr{A}_{m,m} \leq \mathscr{A}_{m,m-1} \leq \ldots \leq \mathscr{A}_{m,2} \leq \mathscr{A}_{m,1} = \sum_{s=1}^{n} p_s f(x_s). \tag{6.1}$$

We define the following functionals by taking the differences of refinement of Jensen's inequality given in (6.1).

$$\Theta_1(f) = \mathscr{A}_{m,r} - f\left(\sum_{s=1}^{n} p_s x_s\right), \quad r = 1,\ldots,m, \tag{6.2}$$
$$\Theta_2(f) = \mathscr{A}_{m,r} - \mathscr{A}_{m,k}, \quad 1 \leq r < k \leq m. \tag{6.3}$$

Under the assumptions of Theorem 6.1, we have

$$\Theta_i(f) \geq 0, \quad i = 1,2. \tag{6.4}$$

Inequalities (6.4) are reversed if f is concave on I.

6.1.1 Inequalities for Csiszár divergence

In [19, 20] Csiszár introduced the following notion.

Definition 6.1 *Let $f : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex function, let $\mathbf{r} = (r_1, \ldots, r_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ be positive probability distributions. Then f-divergence functional is defined by*

$$I_f(\mathbf{r}, \mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{r_i}{q_i}\right). \tag{6.5}$$

And he stated that by defining

$$f(0) := \lim_{x \to 0^+} f(x); \quad 0f\left(\frac{0}{0}\right) := 0; \quad 0f\left(\frac{a}{0}\right) := \lim_{x \to 0^+} xf\left(\frac{a}{0}\right), \quad a > 0, \tag{6.6}$$

we can also use the nonnegative probability distributions as well.

In [21], L. Horv́ath, et al gave the following functional on the based of previous definition.

Definition 6.2 *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be a function, let $\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{R}^n$ and $\mathbf{q} = (q_1, \ldots, q_n) \in (0, \infty)^n$ such that*

$$\frac{r_s}{q_s} \in I, \quad s = 1, \ldots, n.$$

Then they define the sum as $\hat{I}_f(\mathbf{r}, \mathbf{q})$ as

$$\hat{I}_f(\mathbf{r}, \mathbf{q}) := \sum_{s=1}^{n} q_s f\left(\frac{r_s}{q_s}\right). \tag{6.7}$$

We apply Theorem 6.1 to $\hat{I}_f(\mathbf{r}, \mathbf{q})$

Theorem 6.2 *Assume (H_1), let $I \subset \mathbb{R}$ be an interval and let $\mathbf{r} = (r_1, \ldots, r_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ are in $(0, \infty)^n$ such that*

$$\frac{r_s}{q_s} \in I, \quad s = 1, \ldots, n.$$

(i) If $f : I \to \mathbb{R}$ is convex function, then

$$\hat{I}_f(\mathbf{r}, \mathbf{q}) = \sum_{s=1}^{n} q_s f\left(\frac{r_s}{q_s}\right) \geq \ldots \geq \frac{(m-1)!}{(l-1)!} \sum_{(i_1, \ldots, i_l) \in I_l} \eta_{I_m, l}(i_1, \ldots, i_l)$$

$$\left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m, i_j}}\right) f\left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m, i_j}}}{\sum\limits_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m, i_j}}}\right) \geq \ldots \geq f\left(\frac{\sum_{s=1}^{n} r_s}{\sum_{s=1}^{n} q_s}\right) \sum_{s=1}^{n} q_s. \tag{6.8}$$

If f is concave function, then inequality signs in (6.8) *are reversed.*
(ii) *If $f : I \to \mathbb{R}$ is a function such that $x \to xf(x) (x \in I)$ is convex, then*

$$\left(\sum_{s=1}^{n} r_s\right) f\left(\sum_{s=1}^{n} \frac{r_s}{\sum_{s=1}^{n} q_s}\right) \leq \ldots \leq \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l)$$

$$\left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right) f\left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right) \leq \ldots \leq \sum_{s=1}^{n} r_s f\left(\frac{r_s}{q_s}\right) = \hat{I}_{idf}(\mathbf{r},\mathbf{q}) . \quad (6.9)$$

Proof. (i) Consider $p_s = \frac{q_s}{\sum_{s=1}^{n} q_s}$ and $x_s = \frac{r_s}{q_s}$ in Theorem 6.1, we have

$$f\left(\sum_{s=1}^{n} \frac{q_s}{\sum_{s=1}^{n} q_s} \frac{r_s}{q_s}\right) \leq \ldots \leq \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l)$$

$$\left(\sum_{j=1}^{l} \frac{\frac{q_{i_j}}{\sum_{s=1}^{n} q_s}}{\alpha_{I_m,i_j}}\right) f\left(\frac{\sum_{j=1}^{l} \frac{\frac{q_{i_j}}{\sum_{i=1}^{n} q_i}}{\alpha_{I_m,i_j}} \frac{r_{i_j}}{q_{i_j}}}{\sum_{j=1}^{l} \frac{\frac{q_{i_j}}{\sum_{i=1}^{n} q_i}}{\alpha_{I_m,i_j}}}\right) \leq \ldots \leq \sum_{s=1}^{n} \frac{q_s}{\sum_{i=1}^{n} q_s} f\left(\frac{r_s}{q_s}\right). \quad (6.10)$$

And taking the sum $\sum_{s=1}^{n} q_i$ we have (6.8).
(ii) Using $f := idf$ (where "*id*" is the identity function) in Theorem 6.1, we have

$$\sum_{s=1}^{n} p_s x_s f\left(\sum_{s=1}^{n} p_s x_s\right) \leq \ldots \leq \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l)$$

$$\left(\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}} x_{i_j}}{\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}}}\right) f\left(\frac{\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}} x_{i_j}}{\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}}}\right) \leq \ldots \leq \sum_{s=1}^{n} p_s x_s f(x_s). \quad (6.11)$$

Now on using $p_s = \frac{q_s}{\sum_{s=1}^{n} q_s}$ and $x_s = \frac{r_s}{q_s}$, $s = 1,\ldots,n$, we get

$$\sum_{s=1}^{n} \frac{q_s}{\sum_{s=1}^{n} q_s} \frac{r_s}{q_s} f\left(\sum_{s=1}^{n} \frac{q_s}{\sum_{s=1}^{n} q_s} \frac{r_s}{q_s}\right) \leq \ldots \leq \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l)$$

$$\left(\sum_{j=1}^{l} \frac{\frac{q_{i_j}}{\sum_{s=1}^{n} q_s}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{\frac{q_{i_j}}{\sum_{s=1}^{n} q_s}}{\alpha_{I_m,i_j}} \frac{r_{i_j}}{q_{i_j}}}{\sum_{j=1}^{l} \frac{\frac{q_{i_j}}{\sum_{s=1}^{n} q_s}}{\alpha_{I_m,i_j}}}\right) f\left(\frac{\sum_{j=1}^{l} \frac{\frac{q_{i_j}}{\sum_{s=1}^{n} q_s}}{\alpha_{I_m,i_j}} \frac{r_{i_j}}{q_{i_j}}}{\sum_{j=1}^{l} \frac{\frac{q_{i_j}}{\sum_{s=1}^{n} q_s}}{\alpha_{I_m,i_j}}}\right)$$

$$\leq \ldots \leq \sum_{s=1}^{n} \frac{q_s}{\sum_{s=1}^{n} q_s} \frac{r_s}{q_s} f\left(\frac{r_s}{q_s}\right). \quad (6.12)$$

On taking sum $\sum_{s=1}^{n} q_s$ on both sides, we get (6.9). □

6.1.2 Inequalities for Shannon Entropy

Definition 6.3 (SEE [21]) *The Shannon entropy of positive probability distribution* $\mathbf{r} = (r_1, \dots, r_n)$ *is defined by*

$$H(\boldsymbol{r}) := -\sum_{s=1}^{n} r_s \log(r_s). \tag{6.13}$$

Corollary 6.1 *Assume* (H_1).
(i) *If* $\mathbf{q} = (q_1, \dots, q_n) \in (0, \infty)^n$, *and the base of* log *is greater than* 1, *then*

$$\begin{aligned} -\sum_{s=1}^{n} q_s \log(q_s) \le \ldots \le -\frac{(m-1)!}{(l-1)!} \sum_{(i_1,\dots,i_l)\in I_l} \eta_{I_m,l}(i_1, \dots, i_l) \\ \left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \log\left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \le \ldots \le \log\left(\frac{n}{\sum_{s=1}^{n} q_s}\right) \sum_{s=1}^{n} q_s. \end{aligned} \tag{6.14}$$

If the base of log *is between* 0 *and* 1, *then inequality signs in* (6.14) *are reversed.*
(ii) *If* $\mathbf{q} = (q_1, \dots, q_n)$ *is a positive probability distribution and the base of* log *is greater than* 1, *then we have the estimates for the Shannon entropy of* $\mathbf{q}$

$$\begin{aligned} H(\mathbf{q}) \le \ldots \le -\frac{(m-1)!}{(l-1)!} \sum_{(i_1,\dots,i_l)\in I_l} \eta_{I_m,l}(i_1, \dots, i_l) \\ \left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \log\left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \le \ldots \le \log(n). \end{aligned} \tag{6.15}$$

Proof. (i) Using $f := \log$ and $\mathbf{r} = (1, \dots, 1)$ in Theorem 6.2 (i), we get (6.14).
(ii) It is the special case of (i). □

Definition 6.4 (SEE [21]) *The Kullback-Leibler divergence between the positive probability distribution* $\mathbf{r} = (r_1, \dots, r_n)$ *and* $\mathbf{q} = (q_1, \dots, q_n)$ *is defined by*

$$D(\mathbf{r}, \mathbf{q}) := \sum_{s=1}^{n} r_i \log\left(\frac{r_i}{q_i}\right). \tag{6.16}$$

Corollary 6.2 *Assume* (H_1).
(i) *Let* $\mathbf{r} = (r_1, \dots, r_n) \in (0, \infty)^n$ *and* $\mathbf{q} := (q_1, \dots, q_n) \in (0, \infty)^n$. *If the base of* log *is greater than* 1, *then*

$$\begin{aligned} \sum_{s=1}^{n} r_s \log\left(\sum_{s=1}^{n} \frac{r_s}{\sum_{s=1}^{n} q_s}\right) \le \ldots \le \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\dots,i_l)\in I_l} \eta_{I_m,l}(i_1, \dots, i_l) \\ \left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right) \log\left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right) \le \ldots \le \sum_{s=1}^{n} r_s \log\left(\frac{r_s}{q_s}\right). \end{aligned} \tag{6.17}$$

(ii) *If* $\boldsymbol{r}$ *and* $\boldsymbol{q}$ *are positive probability distributions, and the base of* log *is greater than* 1*, then we have*

$$D(\boldsymbol{r},\boldsymbol{q}) \geq \ldots \geq \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l)$$

$$\left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right) \log \left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right) \geq \ldots \geq 0. \tag{6.18}$$

If the base of log *is between* 0 *and* 1*, then inequality signs in* (6.18) *are reversed.*

Proof. (i) On taking $f := \log$ in Theorem 6.2 (ii), we get (6.17).
(ii) It is a special case of (i). □

6.1.3 Inequalities for Rényi Divergence and Entropy

The Rényi divergence and entropy come from [29].

Definition 6.5 *Let* $\boldsymbol{r} := (r_1,\ldots,r_n)$ *and* $\boldsymbol{q} := (q_1,\ldots,q_n)$ *be positive probability distribution, and let* $\lambda \geq 0$, $\lambda \neq 1$.
(a) *The Rényi divergence of order* λ *is defined by*

$$D_\lambda(\boldsymbol{r},\boldsymbol{q}) := \frac{1}{\lambda-1} \log\left(\sum_{i=1}^{n} q_i \left(\frac{r_i}{q_i}\right)^{\lambda}\right). \tag{6.19}$$

(b) *The Rényi entropy of order* λ *of* $\boldsymbol{r}$ *is defined by*

$$H_\lambda(\boldsymbol{r}) := \frac{1}{1-\lambda} \log\left(\sum_{i=1}^{n} r_i^{\lambda}\right). \tag{6.20}$$

The Rényi divergence and the Rényi entropy can also be extended to non-negative probability distributions. If $\lambda \to 1$ in (6.19), we have the Kullback-Leibler divergence, and if $\lambda \to 1$ in (6.20), then we have the Shannon entropy. In the next two results, inequalities can be found for the Rényi divergence.

Theorem 6.3 *Assume* (H_1)*, let* $\boldsymbol{r} = (r_1,\ldots,r_n)$ *and* $\boldsymbol{q} = (q_1,\ldots,q_n)$ *are probability distributions.*
(i) *If* $0 \leq \lambda \leq \mu$ *such that* $\lambda,\mu \neq 1$*, and the base of* log *is greater than* 1*, then*

$$D_\lambda(\boldsymbol{r},\boldsymbol{q}) \leq \ldots \leq$$

$$\frac{1}{\mu-1} \log\left(\frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l) \left(\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum\limits_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}} \left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum\limits_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)^{\frac{\mu-1}{\lambda-1}}\right.$$

$$\leq \ldots \leq D_{\mu}(\boldsymbol{r},\boldsymbol{q}). \tag{6.21}$$

The reverse inequalities hold if the base of log *is between 0 and 1.*
(ii) If $1 < \mu$ *and the base of* log *is greater than 1, then*

$$\begin{aligned} D_1(\boldsymbol{r},\boldsymbol{q}) = D(\boldsymbol{r},\boldsymbol{q}) &= \sum_{s=1}^{n} r_s \log\left(\frac{r_s}{q_s}\right) \\ &\leq \ldots \leq \frac{1}{\mu-1}\log\left(\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\right. \\ &\quad \left.\exp\left(\frac{(\mu-1)\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\log\left(\frac{r_{i_j}}{q_{i_j}}\right)}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)\right) \leq \ldots \leq D_{\mu}(\boldsymbol{r},\boldsymbol{q}), \end{aligned} \tag{6.22}$$

where the base of exp *is same as the base of* log*, and the reverse inequalities hold if the base of* log *is between 0 and 1.*
(iii) If $0 \leq \lambda < 1$*, and the base of* log *is greater than 1, then*

$$\begin{aligned} &D_{\lambda}(\boldsymbol{r},\boldsymbol{q}) \leq \ldots \leq \\ &\frac{1}{\lambda-1}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right) \\ &\leq \ldots \leq D_1(\boldsymbol{r},\boldsymbol{q}). \end{aligned} \tag{6.23}$$

Proof. By applying Theorem 6.1 with $I = (0,\infty)$, $f : (0,\infty) \to \mathbb{R}$, $f(t) := t^{\frac{\mu-1}{\lambda-1}}$

$$p_s := r_s, \quad x_s := \left(\frac{r_s}{q_s}\right)^{\lambda-1}, \ s = 1,\ldots,n,$$

we have

$$\begin{aligned} &\left(\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\right)^{\frac{\mu-1}{\lambda-1}} = \left(\sum_{s=1}^{n} r_s\left(\frac{r_s}{q_s}\right)^{\lambda-1}\right)^{\frac{\mu-1}{\lambda-1}} \\ &\leq \ldots \leq \frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)^{\frac{\mu-1}{\lambda-1}} \\ &\leq \ldots \leq \sum_{s=1}^{n} r_s\left(\left(\frac{r_s}{q_s}\right)^{\lambda-1}\right)^{\frac{\mu-1}{\lambda-1}}, \end{aligned} \tag{6.24}$$

if either $0 \leq \lambda < 1 < \beta$ or $1 < \lambda \leq \mu$, and the reverse inequality in (6.24) holds if $0 \leq \lambda \leq \beta < 1$. By raising to power $\frac{1}{\mu-1}$, we have from all

$$\left(\sum_{s=1}^{n} q_s \left(\frac{r_s}{q_s}\right)^{\lambda}\right)^{\frac{1}{\lambda-1}} \leq \ldots \leq$$

$$\left(\frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l) \left(\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}} \left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)^{\frac{\mu-1}{\lambda-1}}\right)^{\frac{1}{\mu-1}}$$

$$\leq \ldots \leq \left(\sum_{s=1}^{n} r_s \left(\left(\frac{r_s}{q_s}\right)^{\lambda-1}\right)^{\frac{\mu-1}{\lambda-1}}\right)^{\frac{1}{\mu-1}} = \left(\sum_{s=1}^{n} q_s \left(\frac{r_s}{q_s}\right)^{\mu}\right)^{\frac{1}{\mu-1}}. \tag{6.25}$$

Since log is increasing if the base of log is greater than 1, it now follows (6.21). If the base of log is between 0 and 1, then log is decreasing and therefore inequality in (6.21) are reversed. If $\lambda = 1$ and $\beta = 1$, we have (ii) and (iii) respectively by taking limit. □

Theorem 6.4 *Assume (H_1), let $\boldsymbol{r} = (r_1,\ldots,r_n)$ and $\boldsymbol{q} = (q_1,\ldots,q_n)$ are probability distributions. If either $0 \leq \lambda < 1$ and the base of* log *is greater than 1, or $1 < \lambda$ and the base of* log *is between 0 and 1, then*

$$\frac{1}{\sum_{s=1}^{n} q_s \left(\frac{r_s}{q_s}\right)^{\lambda}} \sum_{s=1}^{n} q_s \left(\frac{r_s}{q_s}\right)^{\lambda} \log\left(\frac{r_s}{q_s}\right) \leq \ldots \leq$$

$$\frac{1}{(\lambda-1)\sum_{s=1}^{n} q_s \left(\frac{r_s}{q_s}\right)^{\lambda}} \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l)$$

$$\left(\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}} \left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}\right) \log\left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}} \left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right) \leq \ldots \leq D_{\lambda}(\boldsymbol{r},\boldsymbol{q}) \leq \ldots \leq$$

$$\frac{1}{\lambda-1} \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l) \left(\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}\right) \log\left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}} \left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)$$

$$\leq \ldots \leq D_1(\boldsymbol{r},\boldsymbol{q}). \tag{6.26}$$

The inequalities in (6.26) *are reversed if either $0 \leq \lambda < 1$ and the base of* log *is between 0 and 1, or $1 < \lambda$ and the base of* log *is greater than 1.*

Proof. We prove only the case when $0 \leq \lambda < 1$ and the base of log is greater than 1 and the other cases can be proved similarly. Since $\frac{1}{\lambda-1} < 0$ and the function log is concave then choose $I = (0,\infty)$, $f := \log$, $p_s = r_s$, $x_s := \left(\frac{r_s}{q_s}\right)^{\lambda-1}$ in Theorem 6.1, we have

$$D_\lambda(\mathbf{r},\mathbf{q}) = \frac{1}{\lambda-1}\log\left(\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\right) = \frac{1}{\lambda-1}\log\left(\sum_{s=1}^{n} r_s\left(\frac{r_s}{q_s}\right)^{\lambda-1}\right)$$

$$\leq \dots \leq \frac{1}{\lambda-1}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\dots,i_l)\in I_l}\eta_{I_m,l}(i_1,\dots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\frac{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)$$

$$\leq \dots \leq \frac{1}{\lambda-1}\sum_{s=1}^{n} r_s\log\left(\left(\frac{r_s}{q_s}\right)^{\lambda-1}\right) = \sum_{s=1}^{n} r_s\log\left(\frac{r_s}{q_s}\right) = D_1(\mathbf{r},\mathbf{q}) \tag{6.27}$$

and this give the upper bound for $D_\lambda(\mathbf{r},\mathbf{q})$.
Since the base of log is greater than 1, the function $x \mapsto xf(x)$ $(x > 0)$ is convex therefore $\frac{1}{1-\lambda} < 0$ and Theorem 6.1 gives

$$\begin{aligned}
D_\lambda(\mathbf{r},\mathbf{q}) &= \frac{1}{\lambda-1}\log\left(\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\right) \\
&= \frac{1}{\lambda-1\left(\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\right)}\left(\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\right)\log\left(\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\right) \\
&\geq \dots \geq \frac{1}{\lambda-1\left(\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\right)}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\dots,i_l)\in I_l}\eta_{I_m,l}(i_1,\dots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right) \\
&\left(\frac{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)\log\left(\frac{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right) = \\
&\frac{1}{\lambda-1\left(\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\right)}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\dots,i_l)\in I_l}\eta_{I_m,l}(i_1,\dots,i_l) \\
&\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}\right)\log\left(\frac{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)
\end{aligned}$$

$$\geq \ldots \geq \frac{1}{\lambda-1}\sum_{s=1}^{n} r_s\left(\frac{r_s}{q_s}\right)^{\lambda-1}\log\left(\frac{r_s}{q_s}\right)^{\lambda-1}\frac{1}{\sum_{s=1}^{n} r_s\left(\frac{r_s}{q_s}\right)^{\lambda-1}}$$

$$= \frac{1}{\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}}\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\log\left(\frac{r_s}{q_s}\right) \tag{6.28}$$

which give the lower bound of $D_\lambda(\mathbf{r},\mathbf{q})$. □

By using the previous results, some inequalities are Rényi entropy are obtained. Let $\frac{\mathbf{1}}{\mathbf{n}} = (\frac{1}{n},\ldots,\frac{1}{n})$ be a discrete probability distribution.

Corollary 6.3 *Assume (H_1) and let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{q} = (q_1,\ldots,q_n)$ are positive probability distributions.*
(i) If $0 \leq \lambda \leq \mu$, $\lambda,\mu \neq 1$, and the base of log *is greater than 1, then*

$$H_\lambda(\boldsymbol{p}) \geq \ldots \geq \frac{1}{\mu-1}\log\left(\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\right.$$
$$\left.\times\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}^{\lambda}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)^{\frac{\mu-1}{\lambda-1}}\right) \geq \ldots \geq H_\mu(\boldsymbol{r}). \tag{6.29}$$

The reverse inequalities holds in (6.29) *if the base of* log *is between 0 and 1.*
(ii) If $1 < \mu$ and base of log *is greater than 1, then*

$$H(\boldsymbol{r}) \;= -\sum_{s=1}^{n}\log(p_i) \geq \ldots \geq \log(n) + \frac{1}{1-\mu}\log\left(\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\right.$$
$$\left.\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\exp\left(\frac{(\mu-1)\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\log\left(nr_{i_j}\right)}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)\right) \geq \ldots \geq H_\mu(\boldsymbol{r}), \tag{6.30}$$

where the base of exp *is same as the base of* log*. The inequalities in* (6.30) *are reversed if the base of* log *is between 0 and 1.*
(iii) If $0 \leq \lambda < 1$, and the base of log *is greater than 1, then*

$$H_\lambda(\boldsymbol{r}) \geq \ldots \geq \frac{1}{1-\lambda}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}^{\lambda}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)$$
$$\geq \ldots \geq H(\boldsymbol{r}). \tag{6.31}$$

The inequalities in (6.31) *are reversed if the base of* log *is between 0 and 1.*

Proof. (i) Suppose $\mathbf{q} = \frac{\mathbf{1}}{\mathbf{n}}$ then from (6.19), we have

$$D_\lambda(\mathbf{r},\mathbf{q}) = \frac{1}{\lambda-1}\log\left(\sum_{s=1}^{n} n^{\lambda-1} r_s^\lambda\right) = \log(n) + \frac{1}{\lambda-1}\log\left(\sum_{s=1}^{n} r_s^\lambda\right), \tag{6.32}$$

therefore we have

$$H_\lambda(\mathbf{r}) = \log(n) - D_\lambda(\mathbf{r},\frac{\mathbf{1}}{\mathbf{n}}). \tag{6.33}$$

Now using Theorem 6.3 (i) and (6.33), we get

$$H_\lambda(\mathbf{r}) = \log(n) - D_\lambda\left(\mathbf{r},\frac{\mathbf{1}}{\mathbf{n}}\right) \geq \ldots \geq \log(n) - \frac{1}{\mu-1}$$
$$\log\left(n^{\mu-1}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\times\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}^\lambda}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)^{\frac{\mu-1}{\lambda-1}}\right)$$
$$\geq \ldots \geq \log(n) - D_\mu(\mathbf{r},\mathbf{q}) = H_\mu(\mathbf{r}), \tag{6.34}$$

(ii) and (iii) can be proved similarly. □

Corollary 6.4 *Assume* (H_1) *and let* $\boldsymbol{r} = (r_1,\ldots,r_n)$ *and* $\boldsymbol{q} = (q_1,\ldots,q_n)$ *are positive probability distributions.*
If either $0 \leq \lambda < 1$ *and the base of* log *is greater than 1, or* $1 < \lambda$ *and the base of* log *is between 0 and 1, then*

$$-\frac{1}{\sum_{s=1}^{n} r_s^\lambda}\sum_{s=1}^{n} r_s^\lambda\log(r_s) \geq \ldots \geq \frac{1}{(\lambda-1)\sum_{s=1}^{n} r_s^\lambda}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)$$
$$\left(\sum_{j=1}^{l}\frac{r_{i_j}^\lambda}{\alpha_{I_m,i_j}}\right)\log\left(n^{\lambda-1}\frac{\sum_{j=1}^{l}\frac{r_{i_j}^\lambda}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right) \geq \ldots \geq H_\lambda(\boldsymbol{r})$$
$$\geq \ldots \geq \frac{1}{1-\lambda}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}^\lambda}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)$$
$$\geq \ldots \geq H(\boldsymbol{r}). \tag{6.35}$$

The inequalities in (6.35) *are reversed if either* $0 \leq \lambda < 1$ *and the base of* log *is between 0 and 1, or* $1 < \lambda$ *and the base of* log *is greater than 1.*

Proof. The proof is similar to the Corollary 6.3 by using Theorem 6.4. □

6.2 Inequalities by Using Zipf-Mandelbrot Law

The Zipf-Mandelbrot law is defined as follows (see [26]).

Definition 6.6 *Zipf-Mandelbrot law is a discrete probability distribution depending on three parameters $N \in \{1,2,\ldots,\}, q \in [0,\infty)$ and $t > 0$, and is defined by*

$$f(s;N,q,t) := \frac{1}{(s+q)^t H_{N,q,t}}, \quad s = 1,\ldots,N, \tag{6.36}$$

where

$$H_{N,q,t} = \sum_{k=1}^{N} \frac{1}{(k+q)^t}. \tag{6.37}$$

For $q = 0$, the Zipf-Mandelbrot law becomes Zipf's law.

Conclusion 6.1 *Assume (H_1), let $\boldsymbol{r}$ be a Zipf-Mandelbrot law, by Corollary 6.3 (iii), we get. If $0 \leq \lambda < 1$, and the base of* log *is greater than 1, then*

$$\begin{aligned} H_\lambda(\boldsymbol{r}) = \frac{1}{1-\lambda} \log\left(\frac{1}{H_{N,q,t}^{\lambda}} \sum_{s=1}^{n} \frac{1}{(s+q)^{\lambda s}}\right) \geq \ldots \geq \frac{1}{1-\lambda} \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l) \\ \left(\sum_{j=1}^{l} \frac{1}{\alpha_{I_m,i_j}(i_j+q)H_{N.q,t}}\right) \log\left(\frac{1}{H_{N,q,t}^{\lambda-1}} \frac{\sum\limits_{j=1}^{l} \frac{1}{\alpha_{I_m,i_j}(i_j-q)^{\lambda s}}}{\sum\limits_{j=1}^{l} \frac{1}{\alpha_{I_m,i_j}(i_j-q)^{s}}}\right) \\ \geq \ldots \geq \frac{t}{H_{N,q,t}} \sum_{s=1}^{N} \frac{\log(s+q)}{(s+q)^t} + \log(H_{N,q,t}) = H(\boldsymbol{r}). \end{aligned} \tag{6.38}$$

The inequalities in (6.38) *are reversed if the base of* log *is between 0 and 1.*

Conclusion 6.2 *Assume (H_1), let $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ be the Zipf-Mandelbort law with parameters $N \in \{1,2,\ldots\}$, $q_1,q_2 \in [0,\infty)$ and $s_1,s_2 > 0$, respectively, then from Corollary 6.2 (ii), we have If the base of* log *is greater than 1, then*

$$\begin{aligned} \bar{D}(\boldsymbol{r}_1,\boldsymbol{r}_2) = \sum_{s=1}^{n} \frac{1}{(s+q_1)^{t_1} H_{N,q_1,t_1}} \log\left(\frac{(s+q_2)^{t_2} H_{N,q_2,t_2}}{(s+q_1)^{t_1} H_{N,q_2,t_1}}\right) \\ \geq \ldots \geq \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l)\in I_l} \eta_{I_m,l}(i_1,\ldots,i_l) \left(\sum_{j=1}^{l} \frac{\frac{1}{(i_j+q_2)^{t_2} H_{N,q_2,t_2}}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{\frac{1}{(i_j+q_1)^{t_1} H_{N,q_1,t_1}}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{\frac{1}{(i_j+q_2)^{t_2} H_{N,q_2,t_2}}}{\alpha_{I_m,i_j}}}\right) \end{aligned}$$

$$\log\left(\frac{\sum_{j=1}^{l}\frac{\frac{1}{(i_j+q_1)^{t_1}H_{N,q_1,t_1}}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{\frac{1}{(i_j+q_2)^{t_2}H_{N,q_2,t_2}}}{\alpha_{I_m,i_j}}}\right)\geq\ldots\geq 0. \tag{6.39}$$

The inequalities in (6.39) *are reversed if base of* log *is between 0 and 1.*

Under the assumption of Theorem 6.2 (i), define the functionals as follows.

$$\Theta_3(f) = \mathscr{A}_{m,r}^{[1]} - f\left(\frac{\sum_{s=1}^{n} r_s}{\sum_{s=1}^{n} q_s}\right)\sum_{s=1}^{n} q_s, \quad r=1,\ldots,m, \tag{6.40}$$

$$\Theta_4(f) = \mathscr{A}_{m,r}^{[1]} - \mathscr{A}_{m,k}^{[1]}, \quad 1\leq r<k\leq m. \tag{6.41}$$

where

$$\mathscr{A}_{m,l}^{[1]} = \frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{q_{i_j}}{\alpha_{I_m,i_j}}\right)f\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right).$$

Under the assumption of Theorem 6.2 (ii), define the functional.

$$\Theta_5(f) = \mathscr{A}_{m,r}^{[2]} - \left(\sum_{s=1}^{n} r_s\right)f\left(\frac{\sum_{s=1}^{n} r_s}{\sum_{s=1}^{n} q_s}\right), \quad r=1,\ldots,m, \tag{6.42}$$

$$\Theta_6(f) = \mathscr{A}_{m,r}^{[2]} - \mathscr{A}_{m,k}^{[2]}, \quad 1\leq r<k\leq m. \tag{6.43}$$

where

$$\mathscr{A}_{m,r}^{[2]} = \frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{q_{i_j}}{\alpha_{I_m,i_j}}\right)\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right)f\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l}\frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right).$$

Under the assumption of Corollary 6.1 (i), define the following functional.

$$\Theta_7(f) = A_{m,r}^{[3]} + \sum_{i=1}^{n} q_i\log(q_i), \;\; r=1,\ldots,n \tag{6.44}$$

$$\Theta_8(f) = A_{m,r}^{[3]} - A_{m,k}^{[3]}, \;\; 1\leq r<k\leq m. \tag{6.45}$$

where

$$A_{m,r}^{[3]} = \frac{-(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{q_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\sum_{j=1}^{l}\frac{q_{i_j}}{\alpha_{I_m,i_j}}\right).$$

Under the assumption of Corollary 6.1 (ii), define the following functionals.

$$\Theta_9(f) = A^{[4]}_{m,r} - H(q), \ \ r = 1,\dots,m \tag{6.46}$$

$$\Theta_{10}(f) = A^{[4]}_{m,r} - A^{[4]}_{m,k}, \ \ 1 \le r < k \le m. \tag{6.47}$$

where

$$A^{[4]}_{m,r} = \frac{-(m-1)!}{(l-1)!} \sum_{(i_1,\dots,i_l)\in I_l} \eta_{I_m,l}(i_1,\dots,i_l) \left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \log\left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right).$$

Under the assumption of Corollary 6.2 (i), let us define the functionals.

$$\Theta_{11}(f) = A^{[5]}_{m,r} - \sum_{s=1}^{n} r_s \log\left(\sum_{s=1}^{n} \log \frac{r_n}{\sum_{s=1}^{n} q_s}\right), \ \ r = 1,\dots,m \tag{6.48}$$

$$\Theta_{12}(f) = A^{[5]}_{m,r} - A^{[5]}_{m,k}, \ \ 1 \le r < k \le m. \tag{6.49}$$

where

$$A^{[5]}_{m,r} = \frac{-(m-1)!}{(l-1)!} \sum_{(i_1,\dots,i_l)\in I_l} \eta_{I_m,l}(i_1,\dots,i_l) \left(\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right) \log\left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}{\sum_{j=1}^{l} \frac{q_{i_j}}{\alpha_{I_m,i_j}}}\right).$$

Under the assumption of Theorem 6.3 (i), consider the following functionals.

$$\Theta_{13}(f) = A^{[6]}_{m,r} - D_\lambda(r,q), \ \ r = 1,\dots,m \tag{6.50}$$

$$\Theta_{14}(f) = A^{[6]}_{m,r} - A^{[6]}_{m,k}, \ \ 1 \le r < k \le m. \tag{6.51}$$

where

$$A^{[6]}_{m,r} = \frac{1}{\mu-1} \log\left(\frac{(m-1)!}{(l-1)!} \sum_{(i_1,\dots,i_l)\in I_l} \eta_{I_m,l}(i_1,\dots,i_l) \left(\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}\right) \left(\frac{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}} \left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l} \frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)^{\frac{\mu-1}{\lambda-1}}\right).$$

Under the assumption of Theorem 6.3 (ii), consider the following functionals

$$\Theta_{15}(f) = A^{[7]}_{m,r} - D_1(r,q), \ \ r = 1,\dots,m \tag{6.52}$$

$$\Theta_{16}(f) = A^{[7]}_{m,r} - A^{[7]}_{m,k}, \ \ 1 \le r < k \le m. \tag{6.53}$$

where

$$A_{m,r}^{[7]} = \frac{1}{\mu-1}\log\left(\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\right.$$
$$\left.\exp\left(\frac{(\mu-1)\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\log\left(\frac{r_{i_j}}{q_{i_j}}\right)}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)\right).$$

Under the assumption of Theorem 6.3 (iii), consider the following functionals

$$\Theta_{17}(f) = A_{m,r}^{[8]} - D_\lambda(r,q),\ \ r = 1,\ldots,m \tag{6.54}$$
$$\Theta_{18}(f) = A_{m,r}^{[8]} - A_{m,k}^{[8]},\ \ 1 \le r < k \le m. \tag{6.55}$$

where

$$A_{m,r}^{[8]}=\frac{1}{\lambda-1}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right).$$

Under the assumption of Theorem 6.4 consider the following functionals

$$\Theta_{19}(f) = A_{m,r}^{[9]} - \frac{1}{\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}}\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\right)^{\lambda}\log\left(\frac{r_s}{q_s}\right),\ \ r = 1,\ldots,m \tag{6.56}$$
$$\Theta_{20}(f) = A_{m,r}^{[9]} - A_{m,k}^{[9]},\ \ 1 \le r < k \le m. \tag{6.57}$$
$$\Theta_{21}(f) = D_\lambda(\mathbf{r},\mathbf{q}) - A_{m,r}^{[9]},\quad r = 1,\ldots,m \tag{6.58}$$
$$\Theta_{22}(f) = A_{m,r}^{[10]} - A_{m,r}^{[10]},\ \ 1 \le r < k \le m. \tag{6.59}$$
$$\Theta_{23}(f) = A_{m,r}^{[10]} - A_{m,r}^{[9]},\ \ 1 \le r < k \le m. \tag{6.60}$$
$$\Theta_{24}(f) = A_{m,r}^{[10]} - D_\lambda(\mathbf{r},\mathbf{q}),\ \ r = 1,\ldots,m \tag{6.61}$$
$$\Theta_{25}(f) = D_1(\mathbf{r},\mathbf{q}) - A_{m,r}^{[9]},\ \ r = 1,\ldots,m \tag{6.62}$$
$$\Theta_{26}(f) = D_1(\mathbf{r},\mathbf{q}) - A_{m,r}^{[10]},\ \ r = 1,\ldots,m \tag{6.63}$$
$$\Theta_{27}(f) = D_1(\mathbf{r},\mathbf{q}) - D_\lambda(\mathbf{r},\mathbf{q}) \tag{6.64}$$

where

$$A_{m,r}^{[9]} = \frac{1}{(\lambda-1)\sum_{s=1}^{n} q_s\left(\frac{r_s}{q_s}\lambda\right)}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}\right)$$

$$\log\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right), \tag{6.65}$$

$$A_{m,r}^{[10]} = \frac{1}{\lambda-1}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}}{q_{i_j}}\right)^{\lambda-1}}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right).$$

Under the assumption of Corollary 6.3 (i), consider the following functionals.

$$\Theta_{28}(f) = H_\lambda(\mathbf{p}) - A_{m,r}^{[11]},\ r=1,\ldots,m \tag{6.66}$$

$$\Theta_{29}(f) = A_{m,r}^{[11]} - A_{m,k}^{[11]},\ 1\le r<k\le m. \tag{6.67}$$

$$\Theta_{30}(f) = H_\lambda(\mathbf{p}) - H_\mu(\mathbf{r}) \tag{6.68}$$

$$\Theta_{31}(f) = A_{m,r}^{[11]} - H_\mu(\mathbf{r}),\ r=1,\ldots,m \tag{6.69}$$

where

$$A_{m,r}^{[11]} = \frac{1}{\mu-1}\log\left(\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\left(\frac{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\left(\frac{r_{i_j}^{\lambda}}{\alpha_{I_m,i_j}}\right)}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)^{\frac{\mu-1}{\lambda-1}}\right).$$

Under the assumption of Corollary 6.3 (ii), consider the following functionals

$$\Theta_{32}(f) = H(\mathbf{r}) - A_{m,r}^{[12]},\ r=1,\ldots,m \tag{6.70}$$

$$\Theta_{33}(f) = A_{m,r}^{[12]} - A_{m,k}^{[12]},\ 1\le r<k\le m. \tag{6.71}$$

$$\Theta_{34}(f) = H(\mathbf{r}) - H_\mu(\mathbf{r}) \tag{6.72}$$

$$\Theta_{35}(f) = A_{m,r}^{[12]} - H_\mu(\mathbf{r}),\ r=1,\ldots,m \tag{6.73}$$

where

$$A_{m,r}^{[12]} = \log(n) + \frac{1}{1-\mu}\log\left(\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\exp\left(\frac{(\mu-1)\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\log\left(nr_{i_j}\right)}{\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)\right).$$

Under the assumption of Corollary 6.3 (iii), consider the following functionals.

$$\Theta_{36}(f) = H_{\lambda}(\mathbf{r}) - A^{[13]}_{m,r}, \quad r = 1,\ldots,m \tag{6.74}$$

$$\Theta_{37}(f) = A^{[13]}_{m,r} - A^{[13]}_{m,k}, \quad 1 \le r < k \le m. \tag{6.75}$$

$$\Theta_{38}(f) = H_{\lambda}(\mathbf{r}) - H(\mathbf{r}) \tag{6.76}$$

$$\Theta_{39}(f) = A^{[13]}_{m,r} - H(\mathbf{r}), \quad r = 1,\ldots,m \tag{6.77}$$

where

$$A^{[13]}_{m,r} = \frac{1}{1-\lambda}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\frac{\sum\limits_{j=1}^{l}\frac{r^{\lambda}_{i_j}}{\alpha_{I_m,i_j}}}{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right).$$

Under the assumption of Corollary 6.4, defined the following functionals.

$$\Theta_{40} = -\frac{1}{\sum_{s=1}^{n} r_s^{\lambda}}\sum_{s=1}^{n} r_s^{\lambda}\log(r_s) - A^{[14]}_{m,r} \quad r = 1,\ldots,m \tag{6.78}$$

$$\Theta_{41} = A^{[14]}_{m,k} - A^{[14]}_{m,r}, \quad 1 \le r < k \le m. \tag{6.79}$$

$$\Theta_{42} = A^{[14]}_{m,r} - H_{\lambda}(\mathbf{r}), \quad r = 1,\ldots,m \tag{6.80}$$

$$\Theta_{43} = H_{\lambda}(\mathbf{r}) - A^{[15]}_{m,r}, \quad r = 1,\ldots,m \tag{6.81}$$

$$\Theta_{44} = H_{\lambda}(\mathbf{r}) - H(\mathbf{r}) \tag{6.82}$$

$$\Theta_{45} = A^{[15]}_{m,k} - A^{[15]}_{m,r}, \quad 1 \le r < k \le m. \tag{6.83}$$

$$\Theta_{46} = A^{[15]}_{m,r} - H(\mathbf{r}), \quad r = 1,\ldots,m \tag{6.84}$$

where

$$A^{[14]}_{m,r} = \frac{1}{(\lambda-1)\sum_{s=1}^{n} r_s^{\lambda}}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r^{\lambda}_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(n^{\lambda-1}\frac{\sum\limits_{j=1}^{l}\frac{r^{\lambda}_{i_j}}{\alpha_{I_m,i_j}}}{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right)$$

$$A^{[15]}_{m,r} = \frac{1}{1-\lambda}\frac{(m-1)!}{(l-1)!}\sum_{(i_1,\ldots,i_l)\in I_l}\eta_{I_m,l}(i_1,\ldots,i_l)\left(\sum_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}\right)\log\left(\frac{\sum\limits_{j=1}^{l}\frac{r^{\lambda}_{i_j}}{\alpha_{I_m,i_j}}}{\sum\limits_{j=1}^{l}\frac{r_{i_j}}{\alpha_{I_m,i_j}}}\right).$$

6.3 Generalization of refinement of Jensen's, Rényi and Shannon type inequalities via Green Function

In [6], the green function $G : [\alpha_1, \alpha_2] \times [\alpha_1, \alpha_2] \to \mathbb{R}$ is defined as

$$G(u,v) = \begin{cases} \frac{(u-\alpha_2)(v-\alpha_1)}{\alpha_2-\alpha_1}, & \alpha_1 \leq v \leq u; \\ \frac{(v-\alpha_2)(u-\alpha_1)}{\alpha_2-\alpha_1}, & u \leq v \leq \alpha_2. \end{cases} \tag{6.85}$$

The function G is convex with respect to v and due to symmetry also convex with respect to u. One can also note that G is continuous function. In [32] it is given that any function $f : [\alpha_1, \alpha_2] \to \mathbb{R}$, such that $f \in C^2([\alpha_1, \alpha_2])$ can be written as

$$f(u) = \frac{\alpha_2 - u}{\alpha_2 - \alpha_1} f(\alpha_1) + \frac{u - \alpha_1}{\alpha_2 - \alpha_1} f(\alpha_2) + \int_{\alpha_1}^{\alpha_2} G(u,v) f''(v) dv. \tag{6.86}$$

Theorem 6.5 *Assume* (H_1)*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. If* $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \ldots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*, then the following statements are equivalent.*
(i) *For every continuous convex function* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$

$$f\left(\sum_{s=1}^{n} p_s x_s\right) \leq \mathscr{A}_{m,m} \leq \mathscr{A}_{m,m-1} \leq \ldots \leq \mathscr{A}_{m,2} \leq \mathscr{A}_{m,1} = \sum_{s=1}^{n} p_s f(x_s). \tag{6.87}$$

(ii) *For all* $v \in [\alpha_1, \alpha_2]$

$$G\left(\sum_{i=1}^{n} p_i x_i, v\right) \leq G_{m,m}(I_m, \boldsymbol{x}, \boldsymbol{p}, v, G) \leq G_{m,m-1}(I_m, \boldsymbol{x}, \boldsymbol{p}, v, G)$$

$$\leq \ldots \leq G_{m,2}(I_m, \boldsymbol{x}, \boldsymbol{p}, v, G) \leq G_{m,1}(I_m, \boldsymbol{x}, \boldsymbol{p}, v, G) = \sum_{i=1}^{n} p_i G(x_i, v) \tag{6.88}$$

where

$$G_{m,l}(I_m, \boldsymbol{x}, \boldsymbol{p}, v, G) = \frac{(m-1)!}{(l-1)!} \sum_{(i_1,\ldots,i_l) \in I_l} \eta_{I_m,l}(i_1, \ldots, i_l) \left(\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}}\right) G\left(\frac{\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}} x_{i_j}}{\sum_{j=1}^{l} \frac{p_{i_j}}{\alpha_{I_m,i_j}}}, v\right).$$

Proof. First we see that (i) implies that (ii). Let (i) is valid. As the function $G(\cdot, v)(v \in [\alpha_1, \alpha_2])$ is continuous and convex, so (6.87) holds for the function $G(\cdot, v)$, which is the required reslut.

To prove (ii) implies (i), suppose the function $f : [\alpha_1, \alpha_2] \to \mathbb{R}$, such that f is convex and $f \in C^2[\alpha_1, \alpha_2]$, and for $1 \le r < k \le m$ consider the difference

$$\mathscr{A}_{m,r} - \mathscr{A}_{m,k} = \frac{(m-1)!}{(r-1)!} \sum_{(i_1,\dots,i_r)\in I_r} \eta_{I_m,l}(i_1,\dots,i_r) \left(\sum_{j=1}^{r} \frac{p_{i_j}}{\alpha_{I_m,i_j}} \right) f\left(\frac{\sum_{j=1}^{r} \frac{p_{i_j}}{\alpha_{I_m,i_j}} x_{i_j}}{\sum_{j=1}^{r} \frac{p_{i_j}}{\alpha_{I_m,i_j}}} \right)$$

$$- \frac{(m-1)!}{(k-1)!} \sum_{(i_1,\dots,i_k)\in I_k} \eta_{I_m,l}(i_1,\dots,i_k) \left(\sum_{j=1}^{k} \frac{p_{i_j}}{\alpha_{I_m,i_j}} \right) f\left(\frac{\sum_{j=1}^{k} \frac{p_{i_j}}{\alpha_{I_m,i_j}} x_{i_j}}{\sum_{j=1}^{k} \frac{p_{i_j}}{\alpha_{I_m,i_j}}} \right).$$

Now using the identity (6.86) for the function f on right side of above equation, and after simple calculation we have

$$\mathscr{A}_{m,r} - \mathscr{A}_{m,k} = \int_{\alpha_1}^{\alpha_2} \left(G_{m,r}(I_m, \mathbf{x}, \mathbf{p}, v, G) - G_{m,l}(I_m, \mathbf{x}, \mathbf{p}, v, G) \right) f''(v) dv.$$

As f is convex so $f'' \ge 0$, since (6.88) is valid so the integrand in the integral of is non-negative. So the non negativity of the integral give (6.87). $\square$

Remark 6.1 Assume (H_1), let $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ be a function where $[\alpha_1, \alpha_2] \subset \mathbb{R}$ be an interval. If $x_1, \dots, x_n \in [\alpha_1, \alpha_2]$ and $p_1, \dots, p_n$ are positive real numbers such that $\sum_{i=1}^{n} p_i = 1$.
(i) If for all $v \in [\alpha_1, \alpha_2]$ the inequality (6.88) holds, then

$$\Theta_i(f) \ge 0 \quad i = 1, \dots, 46.$$

(ii) If for all $v \in [\alpha_1, \alpha_2]$ the reverse inequality holds in (6.88), then

$$\Theta_i(f) \le 0 \quad i = 1, \dots, 46.$$

6.4 Generalization of refinement of Jensen's, Rényi and Shannon type inequalities via Montgomery identity

The Montgomery identity via Taylor's formula is given in [2] and [3].

Theorem 6.6 *Let $m \in \mathbb{N}$, $f : I \to \mathbb{R}$ be such that $f^{(m-1)}$ is absolutely continuous, $I \subset \mathbb{R}$ be an open interval $\alpha_1, \alpha_2 \in I$, $\alpha_1 < \alpha_2$. Then the following identity holds*

$$\psi(x) = \frac{1}{\alpha_2 - \alpha_1} \int_{\alpha_1}^{\alpha_2} \psi(u)du + \sum_{k=0}^{m-2} \frac{\psi^{(k+1)}(\alpha_1)(x-\alpha_1)^{k+2}}{k!(k+2)(\alpha_2-\alpha_1)} - \sum_{k=0}^{m-2} \frac{\psi^{(k+1)}(\alpha_2)(x-\alpha_2)^{k+2}}{k!(k+2)(\alpha_2-\alpha_1)}$$

$$+\frac{1}{(m-1)!}\int_{\alpha_1}^{\alpha_2} R_m(x,u)\psi^{(m)}(u)du \tag{6.89}$$

where

$$R_m(x,u)=\begin{cases} -\frac{(x-u)^m}{m(\alpha_2-\alpha_1)}+\frac{x-\alpha_1}{\alpha_2-\alpha_1}(x-u)^{m-1}, & \alpha_1\le u\le x;\\ -\frac{(x-u)^m}{m(\alpha_2-\alpha_1)}+\frac{x-\alpha_2}{\alpha_2-\alpha_1}(x-u)^{m-1}, & x\le u\le \alpha_2. \end{cases} \tag{6.90}$$

Theorem 6.7 *Let $m\in\mathbb{N}$, $f:I\to\mathbb{R}$ be such that $f^{(m-1)}$ is absolutely continuous, $I\subset\mathbb{R}$ be an interval, $\alpha_1,\alpha_2\in I$, $\alpha_1<\alpha_2$. Then the following identity holds*

$$\psi(x)=\frac{1}{\alpha_2-\alpha_1}\int_{\alpha_1}^{\alpha_2}\psi(u)du+\sum_{k=0}^{m-2}\psi^{(k+1)}(x)\frac{(\alpha_1-x)^{k+2}-(\alpha_2-x)^{k+2}}{(k+2)!(\alpha_2-\alpha_1)}$$
$$+\frac{1}{(m-1)!}\int_{\alpha_1}^{\alpha_2}\widehat{R}(x,u)\psi^{(m)}(u)du \tag{6.91}$$

where

$$\widehat{R}(x,u)=\begin{cases} -\frac{1}{m(\alpha_2-\alpha_1)}(\alpha_1-u), & \alpha_1\le u\le x;\\ -\frac{1}{m(\alpha_2-\alpha_1)}(\alpha_2-u), & x\le u\le \alpha_2. \end{cases} \tag{6.92}$$

In case $m=1$, the sum $\sum_{k=0}^{m-2}\ldots$ is empty, so (6.89) and (6.91) reduce to well-known Montgomery identity (see [27])

$$f(x)=\frac{1}{\alpha_2-\alpha_1}\int_{\alpha_1}^{\alpha_2}f(t)dt+\frac{1}{\alpha_2-\alpha_1}\int_{\alpha_1}^{\alpha_2}p(x,u)f'(u)du,$$

where $p(x,u)$ is the Peano kernel, defined by

$$p(x,u)=\begin{cases} \frac{u-\alpha_1}{\alpha_2-\alpha_1}, & \alpha_1\le u\le x;\\ \frac{u-\alpha_2}{\alpha_2-\alpha_1}, & x\le u\le \alpha_2. \end{cases}$$

We construct the identity for the (6.3) with the help of generalized Montgomery identity (6.89).

Theorem 6.8 *Assume (H_1), let $f:[\alpha_1,\alpha_2]\to\mathbb{R}$ be a function where $[\alpha_1,\alpha_2]\subset\mathbb{R}$ be an interval. Also let $x_1,\ldots,x_n\in[\alpha_1,\alpha_2]$ and $p_1,\ldots,p_n$ are positive real numbers such that $\sum_{i=1}^{n}p_i=1$, and $R_m(x,u)$ be the same as defined in* (6.90)*, then the following identity holds.*

$$\Theta_i(f)=\frac{1}{\alpha_2-\alpha_1}\sum_{k=0}^{m-2}\left(\frac{1}{k!(k+2)}\right)\left(f^{(k+1)}(\alpha_1)\Theta_i((x-\alpha_1)^{k+1})-f^{(k+1)}(\alpha_2)\right.$$
$$\left.\times\Theta_i((x-\alpha_2)^{k+1})\right)\frac{1}{(m-1)!}\int_{\alpha_1}^{\alpha_2}\Theta_i(R_m(x,u))f^{(m)}(u)du,,\ i=1,\ldots,46. \tag{6.93}$$

Proof. Using (6.89) in (6.2), (6.3) and (6.40)-(6.84), we get the result. □

Theorem 6.9 *Assume* (H_1)*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \ldots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*, and* $R_m(x,u)$ *be the same as defined in* (6.90)*. Let for* $m \geq 2$

$$\Theta_i(R_m(x,u)) \geq 0 \quad \text{for all } u \in [\alpha_1, \alpha_2] \; i = 1, \ldots, 46.$$

If f *is* m*-convex such that* $f^{(m-1)}$ *is absolutely continuous, then*

$$\Theta_i(f) \geq \frac{1}{\alpha_2 - \alpha_1} \sum_{k=0}^{m-2} \left(\frac{1}{k!(k+2)} \right) \Big(f^{(k+1)}(\alpha_1)\Theta_i((x-\alpha_1)^{k+1}) - f^{(k+1)}(\alpha_2)\Theta_i(x-\alpha_2)^{k+1} \Big) \; i = 1, \ldots, 46. \quad (6.94)$$

Proof. As $f^{(m-1)}$ is absolutely continuous on $[\alpha_1, \alpha_2]$, therefore $f^{(m)}$ exists almost everywhere. As f is m-convex, so $f^{(m)}(u) \geq 0$ for all $u \in [\alpha_1, \alpha_2]$(see [28, p.16]). Hence using Theorem 6.8, we get (6.94). □

Theorem 6.10 *Assume* (H_1)*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \ldots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a convex function.*
(i) *If* $m \geq 2$ *is even, then* (6.94) *holds.*
(ii) *Let the* (6.94) *is valid. If the function*

$$\lambda(x) = \frac{1}{\alpha_2 - \alpha_1} \sum_{l=0}^{m-2} \left(\frac{f^{(l+1)}(\alpha_1)(x-\alpha_1)^{l+2} - f^{(l+1)}(\alpha_2)(x-\alpha_2)^{l+2}}{l!(l+2)} \right)$$

is convex, then the right hand side of (6.94) *is non-negative and*

$$\Theta_i(f) \geq 0 \; i = 1, \ldots, 46.$$

Proof. (i) The function $R_m(\cdot, v)$ is convex (see [12]). Hence for even integer $m \geq 2$

$$\Theta_i(R_m(u,v)) \geq 0,$$

therefore from Theorem 6.9, we have (6.94).
(ii) By using the linearity of $\Theta_i(f)$ we can write the right hand side of (6.94) in the form $\Theta_i(\lambda)$. As λ is supposed to be convex therefore the right hand side of (6.94) is non-negative, so $\Theta_i(f) \geq 0$. □

Theorem 6.11 *Assume* (H_1)*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \ldots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*, and* $\widehat{R}_m(x,u)$ *be the same as defined in* (6.92)*, then the following identity holds.*

$$\Theta_i(f) = \frac{1}{\alpha_2 - \alpha_1} \sum_{k=0}^{m-2} \left(\frac{1}{k!(k+2)} \right) \Big(\Theta_i(f^{(k+1)}(x)(\alpha_1 - x)^{k+1}) - \Theta_i(f^{(k+1)}(x)(\alpha_2 - x)^{k+1} \Big)$$

$$+\frac{1}{(m-1)!}\int_{\alpha_1}^{\alpha_2}\Theta_i(\widehat{R}_m(x,u))f^{(m)}(u)du\ \ i=1,\ldots,46. \tag{6.95}$$

Proof. Using (6.91) in (6.2), (6.3) and (6.40)-(6.63), we get the identity (6.95). □

Theorem 6.12 *Assume (H_1), let $f:[\alpha_1,\alpha_2]\to\mathbb{R}$ be a function where $[\alpha_1,\alpha_2]\subset\mathbb{R}$ be an interval. Also let $x_1,\ldots,x_n\in[\alpha_1,\alpha_2]$ and $p_1,\ldots,p_n$ are positive real numbers such that $\sum_{i=1}^{n}p_i=1$, and $R_m(x,u)$ be the same as define in* (6.92)*. Let for $m\geq 2$*

$$\Theta_i(\widehat{R}_m(x,u))\geq 0\ \ \textit{for all}\ u\in[\alpha_1,\alpha_2],\quad i=1,\ldots,46.$$

If f is m-convex such that $f^{(m-1)}$ is absolutely continuous, then

$$\Theta_i(f)\geq\frac{1}{\alpha_2-\alpha_1}\sum_{k=0}^{m-2}\left(\frac{1}{k!(k+2)}\right)\left(\Theta_i(f^{(k+1)}(x)(\alpha_1-x)^{k+1})-\Theta_i(f^{(k+1)}(x)(\alpha_2-x)^{k+1}\right),$$
$$i=1,\ldots,46. \tag{6.96}$$

Proof. As $f^{(m-1)}$ is absolutely continuous on $[\alpha_1,\alpha_2]$, therefore $f^{(m)}$ exists almost everywhere. As f is m-convex, so $f^{(m)}(u)\geq 0$ for all $u\in[\alpha_1,\alpha_2]$(see [28, p.16]). Hence using Theorem 6.11, we get (6.96). □

Remark 6.2 We can get the similar result as given in Theorem 6.10.

Remark 6.3 We can give related mean value theorems, also construct the new families of m-exponentialy convex functions and Cauchy means related to the functionals Θ_i, $i=1,\ldots,43$ as given in [9].

6.5 Generalization of refinement of Jensen's, Rényi and Shannon type inequalities via Lidstone Polynomial

We generalize the refinement of Jensen's inequality for higher order convex function using Lidstone interpolating polynomial. In [32] Widder give the following result.

Lemma 6.1 *If $g\in C^{\infty}([0,1])$, then*

$$g(u)=\sum_{l=0}^{m-1}\left[g^{(2l)}(0)\mathfrak{F}_l(1-u)+g^{(2l)}(0)\mathfrak{F}_l(t)\right]+\int_0^1 G_m(u,s)g^{(2m)}(s)ds$$

where $\mathfrak{F}_l$ is a polynomial of degree $2l+1$ defined by the relation

$$\mathfrak{F}_0(u)=u,\ \mathfrak{F}_m''(u)=\mathfrak{F}_{m-1}(u),\ \mathfrak{F}_m(0)=\mathfrak{F}_m(1)=0,\ m\geq 1, \tag{6.97}$$

and

$$G_1(u,s) = G(u,s) = \begin{cases} (u-1)s, & \alpha_1 \le s \le u \le \alpha_2; \\ (s-1)u, & \alpha_1 \le u \le s \le \alpha_2, \end{cases}$$

is a homogeneous Green's function of the differential operator $\frac{d^2}{d^2 s}$ *on* $[0,1]$*, and with the successive iterates of* $G(u,s)$

$$G_m(u,s) = \int_0^1 G_1(u,p) G_{m-1}(p,s)dp, \quad m \ge 2.$$

The Lidstone polynomial can be expressed in terms of $G_m(u,s)$ as

$$\mathfrak{F}_m(u) = \int_0^1 G_m(u,s)sds.$$

Lindstone series representation of $g \in C^{2m}[\alpha_1,\alpha_2]$ is given by

$$\begin{aligned} g(u) \;=\; & \sum_{l=0}^{m-1} (\alpha_2-\alpha_1)^{2l} g^{(2l)}(\alpha_1)\mathfrak{F}_l\left(\frac{\alpha_2-u}{\alpha_2-\alpha_1}\right) + \sum_{l=0}^{m-1} (\alpha_2-\alpha_1)^{2l} g^{(2l)}(\alpha_2)\mathfrak{F}_l\left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right) \\ & +(\alpha_2-\alpha_1)^{2l-1} \int_{\alpha_1}^{\alpha_2} G_m\left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}, \frac{t-\alpha_1}{\alpha_2-\alpha_1}\right) g^{(2l)}(t)dt. \end{aligned} \tag{6.98}$$

We construct some new identities the with the help of generalized Lidstone polynomial (6.98).

Theorem 6.13 *Assume* (H_1)*, let* $f : [\alpha_1,\alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1,\alpha_2] \subset \mathbb{R}$ *be an interval such that* $f \in C^{2m}[\alpha_1,\alpha_2]$ *for* $m \ge 1$*. Also let* $x_1,\ldots,x_n \in [\alpha_1,\alpha_2]$ *and* $p_1,\ldots,p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*, and* $\mathfrak{F}_m(t)$ *be the same as define in* (6.97)*, then*

$$\begin{aligned} \Theta_i(f) \;=\; & \sum_{k=1}^{m-1} (\alpha_2-\alpha_1)^{2k} f^{(2k)}(\alpha_1)\Theta_i\left(\mathfrak{F}_l\left(\frac{\alpha_2-x}{\alpha_2-\alpha_1}\right)\right) \\ & + \sum_{k=1}^{m-1} (\alpha_2-\alpha_1)^{2k} f^{(2k)}(\alpha_2)\Theta_i\left(\mathfrak{F}_l\left(\frac{x-\alpha_1}{\alpha_2-\alpha_1}\right)\right) \\ & +(\alpha_2-\alpha_1)^{2k-1} \int_{\alpha_1}^{\alpha_2} \Theta_i\left(G_m\left(\frac{x-\alpha_1}{\alpha_2-\alpha_1}, \frac{t-\alpha_1}{\alpha_2-\alpha_1}\right)\right) f^{(2m)}(t)dt, \\ & \qquad i = 1,2,\ldots,46. \end{aligned} \tag{6.99}$$

Proof. Using (6.98) in place of f in $\Theta_i(f)$, $i = 1,2,\ldots,46$, we get (6.99). □

Theorem 6.14 *Assume* (H_1)*, let* $f : [\alpha_1,\alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1,\alpha_2] \subset \mathbb{R}$ *be an interval such that* $f \in C^{2m}[\alpha_1,\alpha_2]$ *for* $m \ge 1$*. Also let* $x_1,\ldots,x_n \in [\alpha_1,\alpha_2]$ *and* $p_1,\ldots,p_n$

are positive real numbers such that $\sum_{i=1}^{n} p_i = 1$, *and* $\mathfrak{F}_m(t)$ *be the same as define in* (6.97), *let for* $m \geq 1$

$$\Theta_i\left(G_m\left(\frac{x-\alpha_1}{\alpha_2-\alpha_1}, \frac{t-\alpha_1}{\alpha_2-\alpha_1}\right)\right) \geq 0, \text{ for all } t \in [\alpha_1, \alpha_2]. \tag{6.100}$$

If f *is* $2m$*-convex function the we have*

$$\Theta_i(f) \geq \sum_{k=1}^{m-1} (\alpha_2-\alpha_1)^{2k} f^{(2k)}(\alpha_1)\Theta_i\left(\mathfrak{F}_l\left(\frac{\alpha_2-x}{\alpha_2-\alpha_1}\right)\right)$$
$$+\sum_{k=1}^{m-1} (\alpha_2-\alpha_1)^{2k} f^{(2k)}(\alpha_2)\Theta_i\left(\mathfrak{F}_l\left(\frac{x-\alpha_1}{\alpha_2-\alpha_1}\right)\right),$$
$$i = 1, 2, \ldots, 46. \tag{6.101}$$

Proof. Since f is $2m$-convex therefore $f^{(2m)} \geq 0$ for all $x \in [\alpha_1, \alpha_2]$, then by using (6.100) in (6.99) we get the required result. □

Theorem 6.15 *Assume* (H_1), *let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \ldots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$, *also suppose that* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *is* $2m$*-convex then the following results are valid.*
(i) *If* m *is odd integer, then for every* $2m$*-convex function* (6.101) *holds.* (ii) *Suppose* (6.101) *holds, if the function*

$$\lambda(u) = \sum_{l=0}^{m-1} (\alpha_2-\alpha_1)^{2l} g^{(2l)}(\alpha_1)\mathfrak{F}_l\left(\frac{\alpha_2-u}{\alpha_2-\alpha_1}\right) + \sum_{l=0}^{m-1} (\alpha_2-\alpha_1)^{2l} g^{(2l)}(\alpha_2)\mathfrak{F}_l\left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)$$

is convex, then the right hand side of (6.101) *is non-negative and we have*

$$\Theta_i(f) \geq 0, \quad i = 1, 2, \ldots, 46. \tag{6.102}$$

Proof. (i) Note that $G_1(u,s) \leq 0$ for $1 \leq u, s, \leq 1$ and also note that $G_m(u,s) \leq 0$ for odd integer m and $G_m(u,s) \geq 0$ for even integer m. As G_1 is convex function and G_{m-1} is positive for odd integer m, therefore

$$\frac{d^2}{d^2u}(G_m(u,s)) = \int_0^1 \frac{d^2}{d^2u} G_1(u,p)G_{m-1}(p,s)dp \geq 0, \; m \geq 2.$$

This shows that G_m is convex in the first variable u if m is convex. Similarly G_m is concave in the first variable if m is even. Hence if m is odd then

$$\Theta_i\left(G_m\left(\frac{x-\alpha_1}{\alpha_2-\alpha_1}, \frac{t-\alpha_1}{\alpha_2-\alpha_1}\right)\right) \geq 0,$$

therefore (6.102) is valid.
(ii) By using the linearity of $\Theta_i(f)$ we can write the right hand side of (6.101) in the form $\Theta_i(\lambda)$. As λ is supposed to be convex therefore the right hand side of (6.101) is non-negative, so $\Theta_i(f) \geq 0$. □

6.6 Generalization of refinement of Jensen's, Rényi and Shannon type inequalities via Taylor Polynomial

In [9], the following functions are consider to generalized the Popoviciu's inequality, defined as

$$(u-v)_+ = \begin{cases} (u-v), & v \le u; \\ 0, & v > u. \end{cases}$$

The well known Taylor formula is as follows.
Let m be a positive integer and $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ be such that $f^{(m-1)}$ is absolutely continuous, then for all $u \in [\alpha_1, \alpha_2]$ the Taylor's formula at point $c \in [\alpha_1, \alpha_2]$ is

$$f(u) = T_{m-1}(f;c;u) + R_{m-1}(f;c;u),$$

where

$$T_{m-1}(f;c;u) = \sum_{l=0}^{m-1} \frac{f^{(l)}(c)}{l!}(u-c)^l,$$

and the remainder is given by

$$R_{m-1}(f;c;u) = \frac{1}{(m-1)!}\int_c^u f^{(m)}(t)(u-t)^{m-1}dt.$$

The Taylor's formula at point α_1 and α_2 is given by:

$$f(u) = \sum_{l=0}^{m-1} \frac{f^{(l)}(\alpha_1)}{l!}(u-\alpha_1)^l + \frac{1}{(m-1)!}\int_{\alpha_1}^{\alpha_2} f^{(m)}(t)\left((u-t)_+^{m-1}\right)dt. \tag{6.103}$$

$$f(u) = \sum_{l=0}^{m-1} \frac{(-1)^l f^{(l)}(\alpha_2)}{l!}(\alpha_2-u)^l + \frac{(-1)^{m-1}}{(m-1)!}\int_{\alpha_1}^{\alpha_2} f^{(m)}(t)\left((t-u)_+^{m-1}\right)dt \tag{6.104}$$

Theorem 6.16 *Assume (H_1), let $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ be a function where $[\alpha_1, \alpha_2] \subset \mathbb{R}$ be an interval. Also let $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ and $p_1, \ldots, p_n$ are positive real numbers such that $\sum_{i=1}^{n} p_i = 1$. Then we have the following identities:*
(i)

$$\Theta_i(f) = \sum_{l=2}^{m-1} \frac{f^{(l)}(\alpha_1)}{l!}\Theta_i\left((u-\alpha_1)^l\right) + \frac{1}{(m-1)!}\int_{\alpha_1}^{\alpha_2} f^{(m)}(t)\Theta_i\left((u-t)_+^{m-1}\right)dt,$$
$$i = 1, \ldots, 46. \tag{6.105}$$

(ii)

$$\Theta_i(f) = \sum_{l=2}^{m-1} \frac{(-1)^l f^{(l)}(\alpha_2)}{l!} \Theta_i\left((\alpha_2 - u)^l\right) + \frac{(-1)^{m-1}}{(m-1)!} \int_{\alpha_1}^{\alpha_2} f^{(m)}(t) \Theta_i\left((t-u)_+^{m-1}\right) dt,$$
$$i = 1, \ldots, 46. \tag{6.106}$$

Proof. Using (6.103) and (6.104) in (6.3), we get the required result. □

Theorem 6.17 *Assume* (H_1)*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \ldots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*. Let* f *is m-convex function such that* $f^{(m-1)}$ *is absolutely continuous. Then we have the following results:*
(i) If

$$\Theta_i\left((u-t)_+^{m-1}\right) \geq 0, \quad t \in [\alpha_1, \alpha_2], \quad i = 1, \ldots, 46.$$

then

$$\Theta_i(f(u)) \geq \sum_{l=2}^{m-1} \frac{f^{(l)}(\alpha_1)}{l!} \Theta_i\left((u - \alpha_1)^l\right). \tag{6.107}$$

(ii) If

$$(-1)^{m-1} \Theta_i\left((t-u)_+^{m-1}\right) \leq 0 \quad t \in [\alpha_1, \alpha_2], \quad i = 1, \ldots, 46.$$

then

$$\Theta_i(f(u)) \geq \sum_{l=2}^{m-1} \frac{(-1)^l f^{(l)}(\alpha_2)}{l!} \Theta_i\left((\alpha_2 - u)^l\right), \quad i = 1, \ldots, 46. \tag{6.108}$$

Proof. Since $f^{(m-1)}$ is absolutely continuous on $[\alpha_1, \alpha_2]$, $f^{(m)}$ exists almost everywhere. As f is m-convex therefore $f^{(m)}(u) \geq 0$ for all $u \in [\alpha_1, \alpha_2]$. Hence using Theorem 6.16 we obtain (6.107) and (6.108). □

Theorem 6.18 *Assume* (H_1)*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \ldots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*. Then the following results are valid.*

(i) If f *is m-convex, then* (6.107) *holds. Also if* $f^{(l)}(\alpha_1) \geq 0$ *for* $l = 2, \ldots, m-1$*, then the right hand side of* (6.107) *will be non-negative.*
(ii) If m is even and f is m-convex, then (6.108) *holds. Also if* $f^{(l)}(\alpha_1) \leq 0$ *for* $l = 2, \ldots, m-1$ *and* $f^{(l)} \geq 0$ *for* $l = 3, \ldots, m-1$*, then right hand side of* (6.108) *will be non-negative.*
(iii)If m is odd and f is m-convex function then (6.108) *is valid. Also if* $f^{(l)}(\alpha_2) \geq 0$ *for* $l = 2, \ldots, m-1$ *and* $f^{(l)}(\alpha_2) \leq 0$ *for* $l = 2, \ldots, m-2$*, then right hand side of* (6.108) *will be non positive.*

6.7 Generalization of refinement of Jensen's, Rényi and Shannon type inequalities via Hermite Interpolating Polynomial

In [1], the Hermite interpolating polynomial is given as follows.
Let $\alpha_1, \alpha_2 \in \mathbb{R}$ with $\alpha_1 < \alpha_2$, and $\alpha_1 = c_1 < c_2 < \ldots < c_l = \alpha_2 (l \geq 2)$ be the points. For $f \in C^{2m}[\alpha_1, \alpha_2]$ a unique polynomial $\sigma_H^{(i)}(s)$ of degree $(m-1)$ exist and satisfying any of the following conditions:
Hermite Conditions

$$\sigma_H^{(i)}(c_j) = f^{(i)}(c_j);\ 0 \leq i \leq k_j,\ 1 \leq j \leq l,\ \sum_{j=1}^{l} k_j + l = m.$$

It is noted that Hermite conditions include the following particular cases.
Lagrange Conditions $(l = m,\ k_j = 0 \text{ for all } i)$

$$\sigma_L(c_j) = f(c_j),\ 1 \leq j \leq m.$$

Type $(q, m-q)$ **Conditions**$(l = 2,\ 1 \leq q \leq m-1,\ k_1 = q-1,\ k_2 = m-q-1)$

$$\sigma_{(q,m)}^{(i)}(\alpha_1) = f^{(i)}(\alpha_1),\ 0 \leq i \leq q-1$$
$$\sigma_{(q,m)}^{(i)}(\alpha_2) = f^{(i)}(\alpha_2),\ 0 \leq i \leq m-q-1.$$

Two Point Taylor Conditions $(m = 2q,\ l = 2,\ k_1 = k_2 = q-1)$

$$\sigma_{2T}^{(i)}(\alpha_1) = f^{(i)}(\alpha_1),\ f_{2T}^{(i)}(\alpha_2) \ = f^{(i)}(\alpha_2).\ \ 0 \leq i \leq q-1$$

In [1], the following result is given.

Theorem 6.19 *Let $-\infty < \alpha_1 < \alpha_2 < \infty$ and $\alpha_1 < c_1 < c_2 < \ldots < c_l \leq \alpha_2$ $(l \geq 2)$ are the given points and $f \in C^m([\alpha_1, \alpha_2])$. Then we have*

$$f(u) = \sigma_H(u) + R_H(f, u), \tag{6.109}$$

where $\sigma_H(u)$ is the Hermite interpolation polynomial that is

$$\sigma_H(u) = \sum_{j=1}^{l} \sum_{i=0}^{k_j} H_{i_j}(u) f^{(i)}(c_j);$$

the H_{i_j} are the fundamental polynomials of the Hermite basis given as

$$H_{i_j}(u) = \frac{1}{i!}\frac{\omega(u)}{(u-c_j)^{k_j+1-i}}\sum_{k=0}^{k_j-i}\frac{1}{k!}\frac{d^k}{du^k}\left(\frac{(u-c_j)^{k_j+1}}{\omega(u)}\right)\Bigg|_{u=c_j}(u-c_j)^k, \tag{6.110}$$

with

$$\omega(u) = \Pi_{j=1}^{l}(u-c_j)^{k_j+1},$$

and the remainder is given by

$$R_H(f,u) = \int_{\alpha_1}^{\alpha_2} G_{H,m}(u,s)f^{(m)}(s)ds,$$

where $G_{H,m}(u,s)$ is defined by

$$G_{H,m}(u,s) = \begin{cases} \sum_{j=1}^{l}\sum_{i=0}^{k_j}\frac{(c_j-s)^{m-i-1}}{(m-i-1)!}H_{i_j}(u), & s \le u; \\ -\sum_{j=r+1}^{l}\sum_{i=0}^{k_j}\frac{(c_j-s)^{m-i-1}}{(m-i-1)!}H_{i_j}(u), & s \ge u. \end{cases}, \tag{6.111}$$

for all $c_r \le s \le c_{r+1}$; $r = 0,1,\ldots,l$, with $c_0 = \alpha_1$ and $c_{l+1} = \alpha_2$.

Remark 6.4 In particular cases, for Lagrange condition from Theorem 6.19, we have

$$f(u) = \sigma_L(u) + R_L(f,u),$$

where $\sigma_L(u)$ is the Lagrange interpolating polynomial that is

$$\sigma_L(u) = \sum_{j=1}^{m}\sum_{k=1,k\neq j}^{m}\left(\frac{u-c_k}{c_j-c_k}\right)f(c_j),$$

and the remainder $R_L(f,u)$ is given by

$$R_L(f,u) = \int_{\alpha_1}^{\alpha_2} G_L(u,s)f^{(m)}(s)ds,$$

with

$$G_L(u,s) = \frac{1}{(m-1)!}\begin{cases} \sum_{j=1}^{r}(c_j-s)^{m-1}\prod_{k=1,k\neq j}^{m}\left(\frac{u-c_k}{c_j-c_k}\right), & s \le u; \\ -\sum_{j=r+1}^{m}(c_j-s)^{m-1}\prod_{k=1,k\neq j}^{m}\left(\frac{u-c_k}{c_j-c_k}\right), & s \ge u. \end{cases}, \tag{6.112}$$

$c_r \le s \le c_{r+1}$ $r = 1,2,\ldots,m-1$, with $c_1 = \alpha_1$ and $c_m = \alpha_2$,

for type $(q,m-q)$ condition, from Theorem 6.19, we have

$$f(u) = \sigma_{(q,m)}(u) + R_{q,m}(f,u),$$

where $\sigma_{(q,m)}(u)$ is $(q,m-q)$ interpolating that is

$$\sigma_{(q,m)}(u) = \sum_{i=0}^{q-1} \tau_i(u) f^{(i)}(\alpha_1) + \sum_{i=0}^{m-q-1} \eta_i(u) f^{(i)}(\alpha_2),$$

with

$$\tau_i(u) = \frac{1}{i!}(u-\alpha_1)^i \left(\frac{u-\alpha_1}{\alpha_1-\alpha_2}\right)^{m-q} \sum_{k=0}^{q-1-i} \binom{m-q+k-1}{k} \left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)^k \tag{6.113}$$

and

$$\eta_i(u) = \frac{1}{i!}(u-\alpha_1)^i \left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)^{q} \sum_{k=0}^{m-q-1-i} \binom{q+k-1}{k} \left(\frac{u-\alpha_2}{\alpha_2-\alpha_1}\right)^k, \tag{6.114}$$

and the remainder $R_{(q,m)}(f,u)$ is defined as

$$R_{(q,m)}(f,u) = \int_{\alpha_1}^{\alpha_2} G_{q,m}(u,s) f^{(m)}(s)ds,$$

with

$$G_{(q,m)}(u,s) = \begin{cases} \sum_{j=0}^{q-1} \left[\sum_{p=0}^{q-1-j} \binom{m-q+p-1}{p} \left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)^p\right] & \\ \times \frac{(u-\alpha_1)^j(\alpha_1-s)^{m-j-1}}{j!(m-j-1)!} \left(\frac{\alpha_2-u}{\alpha_2-\alpha_1}\right)^{m-q}, & \alpha_1 \le s \le u \le \alpha_2; \\ -\sum_{j=0}^{m-q-1} \left[\sum_{\lambda=0}^{m-q-j-1} \binom{q+\lambda-1}{\lambda} \left(\frac{\alpha_2-u}{\alpha_2-\alpha_1}\right)^\lambda\right] & \\ \times \frac{(u-\alpha_2)^j(\alpha_2-s)^{m-j-1}}{j!(m-j-1)!} \left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)^{q}, & \alpha_1 \le u \le s \le \alpha_2. \end{cases} \tag{6.115}$$

From type Two-point Taylor condition from Theorem 6.19, we have

$$f(u) = \sigma_{2T}(u) + R_{2T}(f,u),$$

where

$$\begin{aligned} \sigma_{2T}(u) &= \sum_{i=0}^{q-1}\sum_{k=0}^{q-1-i} \binom{q+k-1}{k} \left[\frac{(u-\alpha_1)^i}{i!}\left(\frac{u-\alpha_2}{\alpha_1-\alpha_2}\right)^q \left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)^k f^{(i)}(\alpha_1)\right. \\ &\quad \left. - \frac{(u-\alpha_2)^i}{i!}\left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)^q \left(\frac{u-\alpha_1}{\alpha_1-\alpha_2}\right)^k f^{(i)}(\alpha_2)\right] \end{aligned}$$

and the remainder $R_{2T}(f,u)$ is given by

$$R_{2T}(f,u) = \int_{\alpha_1}^{\alpha_2} G_{2T}(u,s) f^{(m)}(s)ds$$

with

$$G_{2T}(u,s) = \begin{cases} \frac{(-1)^q}{(2q-1)!} p^m(u,s) \sum_{j=0}^{q-1} \binom{q-1+j}{j} (u-s)^{q-1-j} \delta^j(u,s), & \alpha_1 \le s \le u \le \alpha_2; \\ \frac{(-1)^q}{(2q-1)!} \delta^m(u,s) \sum_{j=0}^{q-1} \binom{q-1+j}{j} (s-u)^{q-1-j} p^j(u,s), & \alpha_1 \le u \le s \le \alpha_2. \end{cases} \tag{6.116}$$

where $p(u,s) = \frac{(s-\alpha_1)(\alpha_2-u)}{\alpha_2-\alpha_1}$, $\delta(u,s) = p(u,s)$ for all $u,s \in [\alpha_1,\alpha_2]$.
In [6] and [25] the positivity of Green's functions is given as follows.

Lemma 6.2 *For the Green function $G_{H,m}(u,s)$ as defined in (6.111), the following results holds.* (i)

$$\frac{G_{H,m}(u,s)}{\omega(u)} > 0 \quad c_1 \le u \le c_l,\ c_1 \le s \le c_l.$$

(ii)

$$G_{H,m}(u,s) \le \frac{1}{(m-1)!(\alpha_2-\alpha_1)} \left|\omega(u)\right|.$$

(iii)

$$\int_{\alpha_1}^{\alpha_2} G_{H,m}(u,s)ds = \frac{\omega(u)}{m!}.$$

Theorem 6.20 *Assume (H_1), let $f:[\alpha_1,\alpha_2] \to \mathbb{R}$ be a function where $[\alpha_1,\alpha_2] \subset \mathbb{R}$ be an interval. Also let $x_1,\ldots,x_n \in [\alpha_1,\alpha_2]$ and $p_1,\ldots,p_n$ are positive real numbers such that $\sum\limits_{i=1}^{n} p_i = 1$. Also let $\alpha_1 = c_1 < c_2 < \ldots < c_l = \alpha_2$ $(l \ge 2)$ be the points and $f \in C^m([\alpha_1,\alpha_2])$. Moreover H_{i_j} be the fundamental polynomials of Hermite basis and $G_{H,m}$ be the green function as defined by* (6.110) *and* (6.111) *respectively. Then we have the following identity.*

$$\Theta_i(f(u)) = \sum_{j=1}^{l}\sum_{i=0}^{k_j} f^{(i)}(c_j)\Theta_i(H_{i_j}(u)) + \int_{\alpha_1}^{\alpha_2} \Theta_i\left(G_{H,m}(u,s)\right) f^{(m)}(s)ds, \quad i = 1,\ldots,46. \tag{6.117}$$

Proof. Using (6.109) and (6.3) and by the linearity of $\Theta_i(f)$ we get the result. □

Theorem 6.21 *Assume (H_1), let $f:[\alpha_1,\alpha_2] \to \mathbb{R}$ be a function where $[\alpha_1,\alpha_2] \subset \mathbb{R}$ be an interval. Also let $x_1,\ldots,x_n \in [\alpha_1,\alpha_2]$ and $p_1,\ldots,p_n$ are positive real numbers such that $\sum\limits_{i=1}^{n} p_i = 1$. Also let $\alpha_1 = c_1 < c_2 \le \ldots < c_l = \alpha_2$ $(l \ge 2)$ be the points and $f \in C^m([\alpha_1,\alpha_2])$. Moreover H_{i_j} be the fundamental polynomials of Hermite basis and $G_{H,m}$ be the green function as defined by* (6.110) *and* (6.111) *respectively. Assume f be m-convex function and*

$$\Theta_i\left(G_{H,m}(u,s)\right) \ge 0 \quad \textit{for all } s \in [\alpha_1,\alpha_2], \quad i = 1,\ldots,46.,$$

then

$$\Theta_i(f(u)) \ge \sum_{j=1}^{l}\sum_{i=0}^{k_j} f^{(i)}(c_j)\Theta_i(H_{i_j}(u)), \quad i = 1,\ldots,46.$$

Proof. Since f is m-convex therefore $f^{(m)}(u) \geq 0$ for all $u \in [\alpha_1, \alpha_2]$. Hence on applying the Theorem 6.20 we get the result. □

Remark 6.5 If (6.118) is reversed then (6.118) is reversed under the assumption of Theorem 6.21.

By using Lagrange conditions we have the following results.

Corollary 6.5 *Let all the assumption of Theorem 6.20 holds. Let G_L be a green function as defined in* (6.112)*. Also f be m-convex function and*

$$\Theta_i\left(G_L(u,s)\right) \geq 0 \quad \text{for all } s \in [\alpha_1, \alpha_2], \quad i = 1, \ldots, 46.,$$

then

$$\Theta_i(f(u)) \geq \sum_{j=1}^{m} f^{(i)}(c_j)\Theta_i\left(\prod_{k=1,k\neq j}^{m}\left(\frac{u-c_j}{c_j-c_k}\right)\right), \quad i = 1,2,\ldots,46.$$

On using the type $(q, m-q)$ conditions we have the following result.

Corollary 6.6 *Let all the assumption of Theorem 6.20 holds, $G_{(q,m)}$ be a green function as defined in* (6.115) *and τ_i and η_i as defined in* (6.113) *and* (6.114) *respectively. Also let f be m-convex function and*

$$\Theta_i\left(G_{(q,m)}(u,s)\right) \geq 0 \quad \text{for all } s \in [\alpha_1, \alpha_2], \quad i = 1, \ldots, 46.,$$

then

$$\Theta_i(f(u)) \geq \sum_{i=0}^{q-1} f^{(i)}(\alpha_1)\Theta_i(\tau_i(u)) + \sum_{i=0}^{m-q-1} f^{(i)}(\alpha_2)\Theta_i(\eta_i(u)), \quad i = 1, \ldots, 46.$$

By using Two-point Taylor condition we can give the following result.

Corollary 6.7 *Let all the assumption of Theorem 6.20 holds, G_{2T} be a green function as defined in* (6.116)*. Also let f be m-convex function and*

$$\Theta_i\left(G_{2T}(u,s)\right) \geq 0 \quad \text{for all } s \in [\alpha_1, \alpha_2], \quad i = 1, \ldots, 46.,$$

then

$$\begin{aligned}\Theta_i(f(u)) \geq \sum_{i=0}^{q-1}\sum_{k=0}^{q-1-i}\binom{q+k-1}{k}\Bigg[& f^{(i)}(\alpha_1)\Theta_i\left(\frac{(u-\alpha_1)^i}{i!}\left(\frac{u-\alpha_2}{\alpha_1-\alpha_2}\right)^q\left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)^k\right) \\ + & f^{(i)}(\alpha_2)\Theta_i\left(\frac{(u-\alpha_2)^i}{i!}\left(\frac{u-\alpha_1}{\alpha_2-\alpha_1}\right)^q\left(\frac{u-\alpha_2}{\alpha_1-\alpha_2}\right)^k\right)\Bigg], \quad i = 1, \ldots, 46.\end{aligned}$$

Theorem 6.22 *Let all the assumption of Theorem* (6.20) *holds,* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be m-convex function.*
(i) *If* k_j *is odd for each* $j = 2, \dots, l$ *then* (6.118) *holds.*
(ii) *Let* (6.118) *be satisfied and the function*

$$F(u) = \sum_{j=1}^{l} \sum_{i=1}^{k_j} f^{(i)}(c_j) H_{i_j}(u)$$

is convex. Then the right hand side of (6.118) *is non-negative and we have*

$$\Theta_i(f(u)) \geq 0, \quad i = 1, \dots, 46.$$

Proof. (i) Since k_j is odd for all $j = 2, \dots, l$ so we have $\omega(u) \geq 0$, we have $G_{H,m-2}(u,s) \geq 0$, so $G_{H,m}$ is convex, therefore $\Theta_i(G_{H,m}(u,s)) \geq 0$, using Theorem 6.21, we get (6.118).
(ii) Similar to the proof of Theorem 6.10. □

Theorem 6.23 *Assume* (H_1)*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \dots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \dots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*. Also let* $\alpha_1 = c_1 < c_2 < \ldots < c_l = \alpha_2$ $(l \geq 2)$ *be the points and* $f \in C^m([\alpha_1, \alpha_2])$*. Furthermore* H_{i_j}*,* $G_{H,m}$ *and* G *be as defined in* (6.110)*,* (6.111) *and* (6.85) *respectively. Then we have*

$$\Theta_i(f(u)) = \int_{\alpha_1}^{\alpha_2} \Theta_i(G(u,t)) \sum_{j=1}^{l} \sum_{i=0}^{k_j} f^{(i+2)}(c_j) H_{i_j}(t) dt$$
$$+ \int_{\alpha_1}^{\alpha_2} \int_{\alpha_1}^{\alpha_2} \Theta_i(G(u,t)) G_{H,m-2}(t,s) f^{(m)}(s) ds dt, \quad i = 1, 2, \dots, 46. \tag{6.118}$$

Proof. Using (6.86) and (6.3) and following the linearity of $\Theta_i(.)$, we have

$$\Theta_i(f(u)) = \int_{\alpha_1}^{\alpha_2} \Theta_i(G(u,t)) f''(t) dt. \tag{6.119}$$

By Theorem 6.19, $f''(t)$ can be expressed as

$$f''(t) = \sum_{j=1}^{l} \sum_{i=0}^{k_j} H_{i_j}(t) f^{(i+2)}(c_j) + \int_{\alpha_1}^{\alpha_2} G_{H,m-2}(t,s) f^{(m)}(s) ds. \tag{6.120}$$

Using (6.120) in (6.119), we get (6.118). □

Theorem 6.24 *Assume* (H_1)*, let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \dots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \dots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$*. Also let* $\alpha_1 = c_1 < c_2 < \ldots < c_l = \alpha_2$ $(l \geq 2)$ *be the points and* $f \in C^m([\alpha_1, \alpha_2])$*.*

Furthermore H_{i_j}, $G_{H,m}$ *and* G *be as defined in* (6.110), (6.111) *and* (6.85) *respectively. Let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be m-convex function and*

$$\int_{\alpha_1}^{\alpha_2} \Theta_i(G(u,t))G_{H,m-2}(t,s)dt \geq 0 \quad t \in [\alpha_1, \alpha_2], \quad i = 1,2,\ldots,46. \tag{6.121}$$

Then

$$\Theta_i(f(u)) \geq \int_{\alpha_1}^{\alpha_2} \Theta_i\left(G(u,t)\right) \sum_{j=1}^{l} \sum_{i=0}^{k_j} f^{(i+2)}(c_j) H_{i_j}(u) du, \quad i = 1,2,\ldots,46. \tag{6.122}$$

Proof. Since the function f is m-convex therefore $f^{(m)}(u) \geq 0$ for all $u \in [\alpha_1, \alpha_2]$. Hence on applying Theorem 6.23 we obtain (6.122). □

Theorem 6.25 *Assume* (H_1), *let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be a function where* $[\alpha_1, \alpha_2] \subset \mathbb{R}$ *be an interval. Also let* $x_1, \ldots, x_n \in [\alpha_1, \alpha_2]$ *and* $p_1, \ldots, p_n$ *are positive real numbers such that* $\sum_{i=1}^{n} p_i = 1$. *Also let* $\alpha_1 = c_1 < c_2 < \ldots < c_l = \alpha_2$ $(l \geq 2)$ *be the points and* $f \in C^m([\alpha_1, \alpha_2])$.
Let $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be m-convex function then the following holds.*
(i) *If* k_j *is odd for each* $j = 2, \ldots, l$ *then* (6.122) *holds.*
(ii) *Let the inequality* (6.122) *be satisfied*

$$F(.) = \sum_{j=1}^{l} \sum_{i=0}^{k_j} f^{(i+2)}(c_j) H_{i_j}(.) \tag{6.123}$$

is non-negative. Then $\Theta_i(f(u)) \geq 0$, *i = 1, 2, ..., 46.*

Proof. (i) Since $G(u,t)$ is convex and weight are positive, so $\Theta_i\left(G(u,t)\right) \geq 0$. Also as k_j is odd for all $j = 2, \ldots, l$, therefore $\omega(t) \geq 0$ and by using Lemma 6.2 (i), we have $G_{H,m-2}(u,s) \geq 0$ so (6.121) holds. Now using the Theorem 6.24 we have (6.122).
(ii) Using (6.123) in (6.122), we get $\Theta_i(f(u)) \geq 0$. □

For the particular case of Hermite conditions, we can give the following corollaries to above Theorem 6.25. By using type $(q, m-q)$ conditions we give the following results.

Corollary 6.8 *Let* τ_i, η_i *be as defined in* (6.113) *and* (6.114) *respectively. Let* $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ *be m-convex function.*
(i) *If* $m - q$ *is even, then the inequality*

$$\Theta_2(f(u)) \geq \int_{\alpha_1}^{\alpha_2} \Theta_2\left(G(u,t)\right) \left(\sum_{i=0}^{q-1} \tau_i(t) f^{(i+2)}(\alpha_1) + \sum_{i=0}^{m-q-1} \eta_i(t) f^{(i+2)}(\alpha_2) \right) dt, \quad i = 1,2,\ldots,46 \tag{6.124}$$

holds.
(ii) *Let the inequality* (6.124) *be satisfied*

$$F(\cdot) = \sum_{i=0}^{q-1} \tau_i(\cdot) f^{(i+2)}(\alpha_1) + \sum_{i=0}^{m-q-1} \eta_i(\cdot) f^{(i+2)}(\alpha_2)$$

is non-negative. Then $\Theta_i(f(u)) \geq 0$, *i = 1, 2, ..., 46.*

On using two points Taylor conditions we can give the following results.

Corollary 6.9 *Let $f : [\alpha_1, \alpha_2] \to \mathbb{R}$ be m-convex function.*
(i) *If m is even, then*

$$\begin{aligned}\Theta_i(f(u)) \geq \int_{\alpha_1}^{\alpha_2} \Theta_i(G(u,t)) \sum_{i=0}^{q-1} \sum_{k=0}^{q-i-1} \binom{q+k-1}{k} \\ \Bigg[\frac{(t-\alpha_1)^i}{i!} \left(\frac{t-\alpha_2}{\alpha_1-\alpha_2}\right)^q \left(\frac{t-\alpha_1}{\alpha_2-\alpha_1}\right)^k f^{(i+2)}(\alpha_1) \\ + \frac{(t-\alpha_2)^i}{i!} \left(\frac{t-\alpha_1}{\alpha_2-\alpha_1}\right)^q \left(\frac{t-\alpha_2}{\alpha_1-\alpha_2}\right)^k f^{(i+2)}(\alpha_2) \Bigg] dt, \quad i = 1, 2, \ldots, 46.\end{aligned}$$

(ii) *Let the inequality* (6.125) *be satisfied and*

$$\begin{aligned}F(t) = \sum_{i=0}^{q-1} \sum_{k=0}^{q-i-1} \binom{q+k-1}{k} \Bigg[\frac{(t-\alpha_1)^i}{i!} \left(\frac{t-\alpha_2}{\alpha_1-\alpha_2}\right)^q \left(\frac{t-\alpha_1}{\alpha_2-\alpha_1}\right)^k f^{(i+2)}(\alpha_1) \\ + \frac{(t-\alpha_2)^i}{i!} \left(\frac{t-\alpha_1}{\alpha_2-\alpha_1}\right)^q \left(\frac{t-\alpha_2}{\alpha_1-\alpha_2}\right)^k f^{(i+2)}(\alpha_2) \Bigg]\end{aligned}$$

is non-negative. Then $\Theta_i(f(u)) \geq 0$, *i = 1, 2, ..., 46.*

Bibliography

[1] R. P. Agarwal, P. J. Y. Wong, *Error inequalities in polynomial interpolation and their applications*, Kluwer Academic Publishers, Dordrecht/Boston/London (1993).

[2] A. A. Aljinović, J. Pečarić, A. Vukelić, *On some Ostrowski type inequalities via Montgomery identity and Taylors formula II*, Tamkang Journal of Mathematics, 36(4), (2005), 279–301.

[3] A. A. Aljinović, J. Pečarić, *On some Ostrowski type inequalities via Montgomery identity and Taylor's formula*, Tamkang Journal of Mathematics, 36(3), (2005), 199–218.

[4] G. Anderson, Y. Ge, *The size distribution of Chinese cities*, Regional Science and Urban Economics, 35(6), (2005), 756–776.

[5] F. Auerbach, *Das Gesetz der Bevölkerungskonzentration*, Petermanns Geographische Mitteilungen, 59, (1913), 74–76.

[6] P. Beesack, (1962) *On the Greens function of an N-point boundary value problem*, Pacific Journal of Mathematics, 12(3), 801–812.

[7] D. Black, V. Henderson, *Urban evolution in the USA*, Journal of Economic Geography, 3(4), (2003), 343–372.

[8] M. Bosker, S. Brakman, H. Garretsen, M. Schramm, *A century of shocks: the evolution of the German city size distribution 1925–1999*, Regional Science and Urban Economics, 38(4), (2008), 330–347.

[9] S. I. Butt, K. A. Khan, J. Pečarić, *Generaliztion of Popoviciu inequality for higher order convex function via Tayor's polynomial*, Acta Univ. Apulensis Math. Inform, 42, (2015), 181–200.

[10] S. I. Butt, N. Mehmood, J. Pečarić, *New generalizations of Popoviciu type inequalities via new green functions and Finks identity*, Transactions of A. Razmadze Mathematical Institute, 171(3), (2017), 293–303.

[11] S. I. Butt, J. Pečarić, *Popoviciu's Inequality For N-convex Functions*, Lap Lambert Academic Publishing (2016).

[12] S. I. Butt, J. Pečarić, *Weighted Popoviciu type inequalities via generalized Montgomery identities*, Hrvatske akademije znanosti i umjetnosti: Matematicke znanosti, (523=19), (2015), 69–89.

[13] S. I. Butt, K. A. Khan, J. Pečarić, *Popoviciu type inequalities via Hermites polynomial*, Mathematical inequalities & applications, 19(4), (2016), 1309–1318.

[14] L. Horváth, *A method to refine the discrete Jensens inequality for convex and mid-convex functions*, Mathematical and Computer Modelling, 54(9–10), (2011), 2451–2459.

[15] L. Horváth, K. A. Khan, J. Pečarić, *Combinatorial Improvements of Jensens Inequality / Classical and New Refinements of Jensens Inequality with Applications*, Monographs in inequalities 8, Element, Zagreb (2014).

[16] L. Horváth, K. A. Khan, J. Pečarić, *Refinement of Jensen's inequality for operator convex functions*, Advances in inequalities and applications, (2014).

[17] L. Horváth, J. Pečarić, *A refinement of discrete Jensen's inequality*, Math. Inequal. Appl. 14, (2011), 777–791.

[18] Y. M. Ioannides, H. G. Overman, *Zipfs law for cities: an empirical examination*, Regional science and urban economics, 33(2), (2003), 127–137.

[19] I. Csiszár, *Information measures: a critical survey*, in: Tans. 7th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Process and 8th European Meeting of Statist., vol. B, pp. 73–86, Academia Prague.

[20] I. Csiszár, (1967) *Information-type measures of difference of probability distributions and indirect observations*, Stud. Sci. Math. Hungar, **2**, (1978), 299–318.

[21] L. Horváth, D. Pecaric, J. Pečarić, *Estimations of f-and Rényi divergences by using a cyclic refinement of the Jensens inequality*, Bulletin of the Malaysian Mathematical Sciences Society, (2017), 1–14.

[22] J. Pečarić, K. A. Khan, I. Perić, *Generalization of Popoviciu type inequalities for symmetric means generated by convex functions*, J. Math. Comput. Sci., Vol. 4, No. 6, (2014), 1091–1113.

[23] S. Kullback, *Information theory and statistics*, Courier Corporation, (1997).

[24] S. Kullback, R. A. Leibler, *On information and sufficiency*, The annals of mathematical statistics, 22(1), 79–86.

[25] A. Yu. Levin, *Some problems bearing on the oscillation of solution of liear differential equations*, Soviet Math. Dokl. **4**(1963), (1951), 121–124.

[26] N. Lovricevic, Đ. Pecaric, J. Pecaric, *Zipf-Mandelbrot law, f-divergences and the Jensen-type interpolating inequalities*, Journal of inequalities and applications, 2018(1), (2018), 36.

[27] D. S. Mitrinović, J. Pečarić, A. M. Fink, *Classical and new inequalities in Analysis*, Kluwer, The Netherlands, (1993).

[28] J. Pečarić, F. Proschan, Y. L. Tong, *Convex functions, Partial Orderings and Statistical Applications*, Academic Press, New York, (1992).

[29] A. Rényi, *On measure of information and entropy*, in: Proceeding of the Fourth Berkely Symposium on Mathematics, Statistics and Probability, pp. 547–561.

[30] K. T. Rosen, M. Resnick, (1980) *The size distribution of cities: an examination of the Pareto law and primacy*, Journal of Urban Economics, 8(2), (1960), 165–186.

[31] K. T. Soo, *Zipf's Law for cities: a cross-country investigation*, Regional science and urban Economics, 35(3), (2005), 239–263.

[32] D. V. Widder, *Completely convex function and Lidstone series*, Trans. Am. Math. Soc. **51**, (1942), 387–398.

[33] G. K. Zipf, *Human behaviour and the principle of least-effort*, Cambridge MA edn. Reading: Addison-Wesley, (1949).

Chapter 7

Improvements of the inequalities for the f-divergence functional with applications to the Zipf-Mandelbrot law

Saad Ihsan Butt, László Horváth, Đilda Pečarić and Josip Pečarić

Abstract. The Jensen's inequality plays a crucial role to obtain inequalities for divergences between probability distributions. In this chapter, we introduce a new functional, based on the f-divergence functional, and then we obtain some estimates for the new functional, the f-divergence and the Rényi divergence by applying a cyclic refinement of the Jensen's inequality. Some inequalities for Rényi and Shannon entropies are obtained too. Zipf-Mandelbrot law is used to illustrate the results.

2010 AMS Subject Classification. Primary 26A51, 26D15, 94A17.
Key words and phrases. Convex function, f-divergence, inequalities for f-divergence, Jensen's inequality, Shannon entropy.

The research of the second author has been supported by the Hungarian National Research, Development and Innovation Office Grant No. KH130513.

7.1 Introduction

Divergences between probability distributions have been introduced to measure the difference between them. A lot of different type of divergences exist, for example the f-divergence (especially, Kullback–Leibler divergence, Hellinger distance and total variation distance), Rényi divergence, Jensen–Shannon divergence, etc. (see [45] and [51]). There are a lot of papers dealing with inequalities for divergences and entropies, see e.g. [44] and [50] and the references therein. The Jensen's inequality plays a crucial role some of these inequalities.

First we give some recent results on integral and discrete Jensens inequalites. We need the following hypotheses:

(H_1) Let $2 \leq k \leq n$ be integers, and let $p_1, \ldots, p_n$ and $\lambda_1, \ldots, \lambda_k$ represent positive probability distributions.

(H_2) Let C be a convex subset of a real vector space V, and $f : C \to \mathbb{R}$ be a convex function.

(H_3) Let $(X, \mathscr{B}, \mu)$ be a probability space.

Let $l \geq 2$ be a fixed integer. The σ-algebra in X^l generated by the projection mappings $pr_m : X^l \to X$ $(m = 1, \ldots, l)$

$$pr_m(x_1, \ldots, x_l) := x_m$$

is denoted by $\mathscr{B}^l$. μ^l means the product measure on $\mathscr{B}^l$: this measure is uniquely (μ is σ-finite) specified by

$$\mu^l(B_1 \times \ldots \times B_l) := \mu(B_1) \ldots \mu(B_l), \quad B_m \in \mathscr{B}, \quad m = 1, \ldots, l.$$

(H_4) Let g be a μ-integrable function on X taking values in an interval $I \subset \mathbb{R}$.

(H_5) Let f be a convex function on I such that $f \circ g$ is μ-integrable on X.

Under the conditions (H_1) and (H_3-H_5) we define

$$C_{int} = C_{int}(f, g, \mu, \mathbf{p}, \lambda)$$

$$:= \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \int_{X^n} f \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} g(x_{i+j})}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) d\mu^n(x_1, \ldots, x_n), \tag{7.1}$$

and for $t \in [0, 1]$

$$C_{par}(t) = C_{par}(t, f, g, \mu, \mathbf{p}, \lambda) := \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right)$$

$$\cdot \int_{X^n} f \left(t \frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} g(x_{i+j})}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} + (1-t) \int_X g \, d\mu \right) d\mu^n(x_1, \ldots, x_n), \tag{7.2}$$

where $i+j$ means $i+j-n$ in case of $i+j>n$.

Now we state cyclic renements of the discrete and integral form of Jensens inequality introduced in [20] (see also [36]):

Theorem 7.1 *Assume* (H_1) *and* (H_2)*. If* $v_1,\ldots,v_n \in C$*, then*

$$f\left(\sum_{i=1}^{n} p_i v_i\right) \le C_{dis} = C_{dis}(f,\mathbf{v},\mathbf{p},\lambda) \tag{7.3}$$

$$:= \sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right) f\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}v_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) \le \sum_{i=1}^{n} p_i f(v_i)$$

where $i+j$ *means* $i+j-n$ *in case of* $i+j>n$*.*

Theorem 7.2 *Assume* (H_1) *and* (H_3-H_5)*. Then*

$$f\left(\int_X g\,d\mu\right) \le C_{par}(t) \le C_{int} \le \int_X f\circ g\,d\mu, \quad t\in[0,1].$$

To give applications in information theory, we introduce some denitions. The following notion was introduced by Csiszár in [2] and [37].

Definition 7.1 *Let* $f:]0,\infty[\, \to\,]0,\infty[$ *be a convex function, and let* $\mathbf{p} := (p_1,\ldots,p_n)$ *and* $\mathbf{q} := (q_1,\ldots,q_n)$ *be positive probability distributions. The* f*-divergence functional is*

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

It is possible to use nonnegative probability distributions in the f-divergence functional, by defining

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

Based on the previous denition, the following new functional was introduced in [9].

Definition 7.2 *Let* $J\subset\mathbb{R}$ *be an interval, and let* $f: J\to\mathbb{R}$ *be a function. Let* $\mathbf{p} := (p_1,\ldots,p_n)\in\mathbb{R}^n$*, and* $\mathbf{q} := (q_1,\ldots,q_n)\in\,]0,\infty[^n$ *such that*

$$\frac{p_i}{q_i}\in J, \quad i=1,\ldots,n. \tag{7.4}$$

Then let

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

As a special case, Shannon entropy and the measures related to it are frequently applied in fields like population genetics, molecular ecology, information theory, dynamical systems and statistical physics(see [21, 22].

Definition 7.3 *The Shannon entropy of a positive probability distribution* $\mathbf{p} := (p_1, \ldots, p_n)$ *is defined by*

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

One of the most famous distance functions used in information theory [27, 30], mathematical statistics [28, 31, 29] and signal processing [23, 26] is Kullback-Leibler distance. The **Kullback-Leibler** distance [13, 25] between the positive probability distributions $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ is defined by

Definition 7.4 *The Kullback-Leibler divergence between the positive probability distributions* $\mathbf{p} := (p_1, \ldots, p_n)$ *and* $\mathbf{q} := (q_1, \ldots, q_n)$ *is defined by*

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

We shall use the so called Zipf-Mandelbrot law.

Definition 7.5 *Zipf-Mandelbrot law is a discrete probability distribution depends on three parameters* $N \in \{1, 2, \ldots\}$, $q \in [0, \infty[$ *and* $s > 0$, *and it is defined by*

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

where

$$H_{N,q,s} := \sum_{k=1}^{N} \frac{1}{(k+q)^s}.$$

If $q = 0$, *then Zipf–Mandelbrot law becomes Zipf's law.*

Zipf's law is one of the basic laws in information science and bibliometrics. Zipf's law is concerning the frequency of words in the text. We count the number of times each word appears in the text. Words are ranked (r) according to the frequency of occurrence (f). The product of these two numbers is a constant: $r \cdot f = c$.

Apart from the use of this law in bibliometrics and information science, Zipf's law is frequently used in linguistics (see [39], p. 167). In economics and econometrics, this distribution is known as Pareto's law which analyze the distribution of the wealthiest members of the community (see [39], p. 125). These two laws are the same in the mathematical sense, they are only applied in a different context (see [42], p. 294).

The same type of distribution that we have in Zipf's and Pareto's law can be also found in other scientific disciplines, such as: physics, biology, earth and planetary sciences, computer science, demography and the social sciences. For example, the same type of distribution, which we also call the Power law, we can analyze the number of hits on web sites, the magnitude of earthquakes, diameter of moon craters, intensity of solar flares, intensity of wars, population of cities, and others (see [48]).

More general model introduced Benoit Mandelbrot (see [46]), by using arguments on the fractal structure of lexical trees.

The are also quite different interpretation of Zipf-Mandelbrot law in ecology, as it is pointed out in [47] (see also [43] and [52]).

7.2 Estimations of f- and Rényi divergences

In this section we obtain some estimates for the new functional, the f-divergence functional, the Sannon entropy and the Rényi divergence by applying cyclic renement results for the Jensens inequality. Finally, some concrete cases are considered, by using Zipf-Mandelbrot law.

It is generally common to take log with base of 2 in the introduced notions, but in our investigations this is not essential.

7.2.1 Inequalities for Csiszár divergence and Shannon entropy

In the first result we apply Theorem 7.1 to $\hat{I}_f(\mathbf{p},\mathbf{q})$.

Theorem 7.3 *Let $2 \le k \le n$ be integers, and let $\lambda := (\lambda_1,\ldots,\lambda_k)$ be a positive probability distribution. Let $J \subset \mathbb{R}$ be an interval, let $\mathbf{p} := (p_1,\ldots,p_n) \in \mathbb{R}^n$, and let $\mathbf{q} := (q_1,\ldots,q_n) \in]0,\infty[^n$ such that*

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

(a) If $f : J \to \mathbb{R}$ is a convex function, then

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

$$\geq \sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) f\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right) \geq f\left(\frac{\sum\limits_{i=1}^{n}p_i}{\sum\limits_{i=1}^{n}q_i}\right)\sum_{i=1}^{n}q_i. \tag{7.5}$$

If f is a concave function, then inequality signs in (7.5) are reversed.

(b) If $f : J \to \mathbb{R}$ is a function such that $x \to xf(x)$ $(x \in J)$ is convex, then

$$\hat{I}_{id_J f}(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} p_i f\left(\frac{p_i}{q_i}\right)$$

$$\geq \sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right) f\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right) \geq f\left(\frac{\sum\limits_{i=1}^{n}p_i}{\sum\limits_{i=1}^{n}q_i}\right)\sum_{i=1}^{n}p_i. \tag{7.6}$$

If $x \to xf(x)$ $(x \in J)$ is a concave function, then inequality signs in (7.6) are reversed. In all these inequalities $i+j$ means $i+j-n$ in case of $i+j>n$.

Proof. (a) By applying Theorem 7.1 with $C := J$, $f := f$,

$$p_i := \frac{q_i}{\sum\limits_{i=1}^{n}q_i}, \quad v_i := \frac{p_i}{q_i}, \quad i=1,\ldots,n$$

we have

$$\sum_{i=1}^{n}q_i f\left(\frac{p_i}{q_i}\right) = \left(\sum_{i=1}^{n}q_i\right)\cdot\sum_{i=1}^{n}\frac{q_i}{\sum\limits_{i=1}^{n}q_i} f\left(\frac{p_i}{q_i}\right)$$

$$\geq \left(\sum_{i=1}^{n}q_i\right)\cdot\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}\frac{q_{i+j}}{\sum\limits_{i=1}^{n}q_i}\right) f\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}\frac{q_{i+j}}{\sum\limits_{i=1}^{n}q_i}\frac{p_{i+j}}{q_{i+j}}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}\frac{q_{i+j}}{\sum\limits_{i=1}^{n}q_i}}\right)$$

$$= \sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) f\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)$$

$$\geq f\left(\frac{\sum\limits_{i=1}^{n}p_i}{\sum\limits_{i=1}^{n}q_i}\right)\sum_{i=1}^{n}q_i.$$

(b) We can prove similarly to (a), by using $f := id_J f$.
The proof is complete. □

Remark 7.1 (a) Csiszár and Körner classical inequality for the f-divergence functional is generalized and refined in (7.5).

(b) Other type of refinements are applied to the f-divergence functional in [40], [41] and [35].

(c) For example, the functions $x \to x\log_b(x)$ $(x>0,\ b>1)$ and $x \to x\arctan(x)$ $(x \in \mathbb{R})$ are convex.

We mention two special cases of the previous result.

The first case corresponds to the entropy of a discrete probability distribution.

Corollary 7.1 *Let $2 \leq k \leq n$ be integers, and let $\lambda := (\lambda_1, \ldots, \lambda_k)$ be a positive probability distribution.*

(a) If $\mathbf{q} := (q_1, \ldots, q_n) \in \,]0,\infty[^n$, and the base of $\log$ is greater than 1, then

$$-\sum_{i=1}^{n} q_i \log(q_i)$$

$$\leq -\sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j}\right) \log\left(\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j}\right) \leq \log\left(\frac{n}{\sum_{i=1}^{n} q_i}\right) \sum_{i=1}^{n} q_i. \tag{7.7}$$

If the base of $\log$ is between 0 and 1, then inequality signs in (7.7) are reversed.

(b) If $\mathbf{q} := (q_1, \ldots, q_n)$ is a positive probability distribution and the base of $\log$ is greater than 1, then we have estimates for the Shannon entropy of $\mathbf{q}$

$$H(\mathbf{q}) \leq -\sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j}\right) \log\left(\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j}\right) \leq \log(n).$$

If the base of $\log$ is between 0 and 1, then inequality signs in (7.7) are reversed.

In all these inequalities $i+j$ means $i+j-n$ in case of $i+j > n$.

Proof. (a) It follows from Theorem 7.3 (a), by using $f := \log$ and $\mathbf{p} := (1, \ldots, 1)$.

(b) It is a special case of (a). □

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

Corollary 7.2 *Let $2 \leq k \leq n$ be integers, and let $\lambda := (\lambda_1, \ldots, \lambda_k)$ be a positive probability distribution.*

(a) Let $\mathbf{p} := (p_1, \ldots, p_n) \in \,]0,\infty[^n$ and $\mathbf{q} := (q_1, \ldots, q_n) \in \,]0,\infty[^n$. If the base of $\log$ is greater than 1, then

$$\sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right) \tag{7.8}$$

$$\geq \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}\right) \log\left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}}{\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j}}\right) \geq \log\left(\frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i}\right) \sum_{i=1}^{n} p_i. \tag{7.9}$$

If the base of $\log$ is between 0 and 1, then inequality signs in (7.9) are reversed.

(b) If $\mathbf{p}$ *and* $\mathbf{q}$ *are positive probability distributions, and the base of* log *is greater than* 1*, then we have*

$$D(\mathbf{p}\|\mathbf{q}) \geq \sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right) \geq 0. \tag{7.10}$$

If the base of log *is between* 0 *and* 1*, then inequality signs in (7.10) are reversed.*

In all these inequalities $i+j$ *means* $i+j-n$ *in case of* $i+j>n$*.*

Proof. (a) We can apply Theorem 7.3 (b) to the function $f := \log$.

(b) It is a special case of (a). □

Remark 7.2 We can apply Theorem 7.3 to have similar inequalities for other distances between two probability distributions.

7.2.2 Inequalities for Rényi divergence and entropy

The Rényi divergence and entropy come from [49].

Definition 7.6 *Let* $\mathbf{p} := (p_1,\ldots,p_n)$ *and* $\mathbf{q} := (q_1,\ldots,q_n)$ *be positive probability distributions, and let* $\alpha \geq 0$, $\alpha \neq 1$*.*

(a) The Rényi divergence of order α *is defined by*

$$D_\alpha(\mathbf{p},\mathbf{q}) := \frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n}q_i\left(\frac{p_i}{q_i}\right)^{\alpha}\right). \tag{7.11}$$

(b) The Rényi entropy of order α *of* $\mathbf{p}$ *is defined by*

$$H_\alpha(\mathbf{p}) := \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{n}p_i^{\alpha}\right). \tag{7.12}$$

The Rényi divergence and the Rényi entropy can also be extended to nonnegative probability distributions.

If $\alpha \to 1$ in (7.11), we have the Kullback-Leibler divergence, and if $\alpha \to 1$ in (7.12), then we have the Shannon entropy.

In the next two results inequalities can be found for the Rényi divergence.

Theorem 7.4 *Let* $2 \leq k \leq n$ *be integers, and let* $\lambda := (\lambda_1,\ldots,\lambda_k)$, $\mathbf{p} := (p_1,\ldots,p_n)$ *and* $\mathbf{q} := (q_1,\ldots,q_n)$ *be positive probability distributions.*

(a) If $0 \leq \alpha \leq \beta$, α, $\beta \neq 1$*, and the base of* log *is greater than* 1*, then*

$$D_\alpha(\mathbf{p},\mathbf{q}) \leq \frac{1}{\beta-1}\log\left(\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)^{\frac{\beta-1}{\alpha-1}}\right. \tag{7.13}$$

$$\leq D_\beta(\mathbf{p},\mathbf{q})$$

The reverse inequalities hold if the base of log *is between* 0 *and* 1.

(b) If $1<\beta$*, and the base of* log *is greater than* 1*, then*

$$D_1(\mathbf{p},\mathbf{q}) = D(\mathbf{p}\|\mathbf{q}) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right)$$

$$\leq \frac{1}{\beta-1}\log\left(\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\exp\left(\frac{(\beta-1)\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\log\left(\frac{p_{i+j}}{q_{i+j}}\right)}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)\right)$$

$$\leq D_\beta(\mathbf{p},\mathbf{q}),$$

where the base of exp *is the same as the base of* log.

The reverse inequalities hold if the base of log *is between* 0 *and* 1.

(c) If $0\leq\alpha<1$*, and the base of* log *is greater than* 1*, then*

$$D_\alpha(\mathbf{p},\mathbf{q})$$

$$\leq \frac{1}{\alpha-1}\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)\leq D_1(\mathbf{p},\mathbf{q})$$

The reverse inequalities hold if the base of log *is between* 0 *and* 1.

In all these inequalities $i+j$ *means* $i+j-n$ *in case of* $i+j>n$.

Proof. (a) By applying Theorem 7.1 with $C :=]0,\infty[$, $f :]0,\infty[\,\to\mathbb{R}$, $f(t) := t^{\frac{\beta-1}{\alpha-1}}$,

$$v_i := \left(\frac{p_i}{q_i}\right)^{\alpha-1}, \quad i=1,\ldots,n,$$

we have

$$\left(\sum_{i=1}^{n} q_i\left(\frac{p_i}{q_i}\right)^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}} = \left(\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right)^{\frac{\beta-1}{\alpha-1}}$$

$$\leq \sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)^{\frac{\beta-1}{\alpha-1}} \leq \sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\beta-1} \tag{7.14}$$

if either $0\leq\alpha<1<\beta$ or $1<\alpha\leq\beta$, and the reverse inequalities hold in (7.61) if $0\leq\alpha\leq\beta<1$. By raising the power $\frac{1}{\beta-1}$, we have from all these cases that

$$\left(\sum_{i=1}^{n} q_i\left(\frac{p_i}{q_i}\right)^{\alpha}\right)^{\frac{1}{\alpha-1}}$$

$$\leq \left(\sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \left(\frac{\sum\limits_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}} \right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right)^{\frac{\beta-1}{\alpha-1}} \right)^{\frac{1}{\beta-1}}$$

$$\leq \left(\sum_{i=1}^{n} p_i \left(\frac{p_i}{q_i} \right)^{\beta-1} \right)^{\frac{1}{\beta-1}} = \left(\sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} \right)^{\beta} \right)^{\frac{1}{\beta-1}}.$$

Since log is increasing if the base of log is greater than 1, it now follows (7.13).

If the base of log is between 0 and 1, then log is decreasing, and therefore inequality signs in (7.13) are reversed.

(b) and (c) When $\alpha = 1$ or $\beta = 1$, we have the result by taking limit.

The proof is complete. □

Theorem 7.5 *Let $2 \leq k \leq n$ be integers, and let $\lambda := (\lambda_1, \ldots, \lambda_k)$, $\mathbf{p} := (p_1, \ldots, p_n)$ and $\mathbf{q} := (q_1, \ldots, q_n)$ be positive probability distributions.*

If either $0 \leq \alpha < 1$ and the base of log *is greater than* 1*, or $1 < \alpha$ and the base of* log *is between* 0 *and* 1*, then*

$$\frac{1}{\sum\limits_{i=1}^{n} q_i \left(\frac{p_i}{q_i} \right)^{\alpha}} \sum_{i=1}^{n} p_i \left(\frac{p_i}{q_i} \right)^{\alpha-1} \log \left(\frac{p_i}{q_i} \right) \leq \frac{1}{(\alpha-1) \sum\limits_{i=1}^{n} p_i \left(\frac{p_i}{q_i} \right)^{\alpha-1}} \times$$

$$\times \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}} \right)^{\alpha-1} \right) \log \left(\frac{\sum\limits_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}} \right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) \leq D_\alpha(\mathbf{p}, \mathbf{q}) \quad (7.15)$$

$$\leq \frac{1}{\alpha-1} \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \log \left(\frac{\sum\limits_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}} \right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) \leq D_1(\mathbf{p}, \mathbf{q})$$

If either $0 \leq \alpha < 1$ and the base of log *is between* 0 *and* 1*, or $1 < \alpha$ and the base of* log *is greater than* 1*, then the reverse inequalities holds.*

In all these inequalities $i + j$ means $i + j - n$ in case of $i + j > n$.

Proof. We prove only the case when $0 \leq \alpha < 1$ and the base of log is greater than 1, the other cases can be proved similarly.

Since $\frac{1}{\alpha-1} < 0$ and the function log is concave, we have from Theorem 7.1 by choosing $C :=]0, \infty[$, $f := \log$,

$$v_i := \left(\frac{p_i}{q_i} \right)^{\alpha-1}, \quad i = 1, \ldots, n,$$

that

$$
\begin{aligned}
D_\alpha(\mathbf{p},\mathbf{q}) &= \frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right) \\
&\leq \frac{1}{\alpha-1}\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) \\
&\leq \frac{1}{\alpha-1}\sum_{i=1}^{n} p_i\log\left(\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right) = \sum_{i=1}^{n} p_i\log\left(\frac{p_i}{q_i}\right) = D_1(\mathbf{p},\mathbf{q})
\end{aligned}
$$

and this gives the desired upper bound for $D_\alpha(\mathbf{p},\mathbf{q})$.

Since the base of log is greater than 1, the function $x \to x\log(x)$ $(x>0)$ is convex, and therefore $\frac{1}{1-\alpha} < 0$ and Theorem 7.1 imply that

$$
\begin{aligned}
D_\alpha(\mathbf{p},\mathbf{q}) &:= \frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right) \\
&= \frac{1}{(\alpha-1)\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\left(\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right)\log\left(\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right) \\
&\geq \frac{1}{(\alpha-1)\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\times \\
&\times\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) \\
&\frac{1}{(\alpha-1)\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) \\
&\geq \frac{1}{(\alpha-1)\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\log\left(\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right) \\
&= \frac{1}{\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\log\left(\frac{p_i}{q_i}\right)
\end{aligned}
$$

which gives the desired lower bound for $D_\alpha(\mathbf{p},\mathbf{q})$.

The proof is complete. □

Now, by using the previous theorems, some inequalities of Rényi entropy are obtained. Denote $\frac{\mathbf{1}}{\mathbf{n}} := \left(\frac{1}{n},\dots,\frac{1}{n}\right)$ be the discrete uniform distribution.

Corollary 7.3 *Let $2 \le k \le n$ be integers, and let $\lambda := (\lambda_1,\dots,\lambda_k)$ and $\mathbf{p} := (p_1,\dots,p_n)$ be positive probability distributions.*

(a) If $0 \le \alpha \le \beta$, α, $\beta \ne 1$, and the base of log *is greater than* 1*, then*

$$H_\alpha(\mathbf{p}) \ge \frac{1}{1-\beta}\log\left(\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}^{\alpha}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)^{\frac{\beta-1}{\alpha-1}}\right) \ge H_\beta(\mathbf{p}).$$

The reverse inequalities hold if the base of log *is between* 0 *and* 1*.*

(b) If $1 < \beta$, and the base of log *is greater than* 1*, then*

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

$$+\frac{1}{1-\beta}\log\left(\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\exp\left(\frac{(\beta-1)\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\log(np_{i+j})}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)\right)$$

$$\ge H_\beta(\mathbf{p}),$$

where the base of exp *is the same as the base of* log*.*

The reverse inequalities hold if the base of log *is between* 0 *and* 1*.*

(c) If $0 \le \alpha < 1$, and the base of log *is greater than* 1*, then*

$$H_\alpha(\mathbf{p}) \ge \frac{1}{1-\alpha}\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}^{\alpha}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) \ge H(\mathbf{p})$$

The reverse inequalities hold if the base of log *is between* 0 *and* 1*.*

In all these inequalities $i+j$ means $i+j-n$ in case of $i+j>n$.

Proof. If $\mathbf{q} = \frac{\mathbf{1}}{\mathbf{n}}$, then

$$D_\alpha(\mathbf{p},\frac{\mathbf{1}}{\mathbf{n}}) = \frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n}n^{\alpha-1}p_i^{\alpha}\right) = \log(n) + \frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n}p_i^{\alpha}\right),$$

and therefore

$$H_\alpha(\mathbf{p}) = \log(n) - D_\alpha(\mathbf{p}, \frac{\mathbf{1}}{\mathbf{n}}). \tag{7.16}$$

(a) It follows from Theorem 7.4 and (7.16) that

$$H_\alpha(\mathbf{p}) = \log(n) - D_\alpha(\mathbf{p}, \frac{\mathbf{1}}{\mathbf{n}})$$

$$\geq \log(n) - \frac{1}{\beta-1} \log\left(n^{\beta-1} \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}^{\alpha}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right)^{\frac{\beta-1}{\alpha-1}} \right)$$

$$\geq \log(n) - D_\beta(\mathbf{p}, \frac{\mathbf{1}}{\mathbf{n}}) = H_\beta(\mathbf{p}).$$

(b) and (c) can be proved similarly.

The proof is complete. □

Corollary 7.4 *Let $2 \leq k \leq n$ be integers, and let $\lambda := (\lambda_1, \ldots, \lambda_k)$ and $\mathbf{p} := (p_1, \ldots, p_n)$ be positive probability distributions.*

If either $0 \leq \alpha < 1$ and the base of $\log$ is greater than 1, or $1 < \alpha$ and the base of $\log$ is between 0 and 1, then

$$-\frac{1}{\sum_{i=1}^{n} p_i^{\alpha}} \sum_{i=1}^{n} p_i^{\alpha} \log(p_i) \geq \log(n) - \frac{1}{(\alpha-1)\sum_{i=1}^{n} p_i^{\alpha}} \times$$

$$\times \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}^{\alpha} \right) \log\left(n^{\alpha-1} \frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}^{\alpha}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) \geq H_\alpha(\mathbf{p})$$

$$\geq \frac{1}{1-\alpha} \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \log\left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}^{\alpha}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) \geq H(\mathbf{p})$$

If either $0 \leq \alpha < 1$ and the base of $\log$ is between 0 and 1, or $1 < \alpha$ and the base of $\log$ is greater than 1, then the reverse inequalities holds.

In all these inequalities $i+j$ means $i+j-n$ in case of $i+j > n$.

Proof. We can prove as Corollary 7.3, by using Theorem 7.5. □

We illustrate our results by using Zipf–Mandelbrot law.

7.2.3 Inequalities by using the Zipf-Mandelbrot law

We illustrate the previous results by using Zipf-Mandelbrot law.

Corollary 7.5 *Let $\mathbf{p}$ be the Zipf-Mandelbrot law as in Definition 10.1, let $2 \le k \le N$ be integers, and let $\lambda := (\lambda_1, \dots, \lambda_k)$ be a probability distribution. By applying Corollary 7.3 (c), we have:*

If $0 \le \alpha < 1$, and the base of $\log$ is greater than 1, then

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log\left(\frac{1}{H^{\alpha}_{N,q,s}} \sum_{i=1}^{N} \frac{1}{(i+q)^{\alpha s}}\right)$$

$$\ge \frac{1}{1-\alpha} \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+j+q)^s H_{N,q,s}}\right) \log\left(\frac{1}{H^{\alpha-1}_{N,q,s}} \frac{\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+j+q)^{\alpha s}}}{\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+q)^{s}}}\right)$$

$$\ge \frac{s}{H_{N,q,s}} \sum_{i=1}^{N} \frac{\log(i+q)}{(i+q)^s} + \log(H_{N,q,s}) = H(\mathbf{p})$$

The reverse inequalities hold if the base of $\log$ is between 0 and 1.

In all these inequalities $i+j$ means $i+j-n$ in case of $i+j>n$.

Corollary 7.6 *Let $\mathbf{p}_1$ and $\mathbf{p}_2$ be the Zipf-Mandelbrot law with parameters $N \in \{1,2,\dots\}$, $q_1,\ q_2 \in [0,\infty[$ and $s_1,\ s_2 > 0$, respectively, let $2 \le k \le N$ be integers, and let $\lambda := (\lambda_1, \dots, \lambda_k)$ be a probability distribution. By applying Corollary 7.2 (b), we have:*

If the base of $\log$ is greater than 1, then

$$D(\mathbf{p}_1 \| \mathbf{p}_2) = \sum_{i=1}^{N} \frac{1}{(i+q_1)^{s_1} H_{N,q_1,s_1}} \log\left(\frac{(i+q_2)^{s_2} H_{N,q_2,s_2}}{(i+q_1)^{s_1} H_{N,q_1,s_1}}\right)$$

$$\ge \sum_{i=1}^{N} \left(\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+q_1)^{s_1} H_{N,q_1,s_1}}\right) \log\left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+q_1)^{s_1} H_{N,q_1,s_1}}}{\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+q_2)^{s_2} H_{N,q_2,s_2}}}\right) \ge 0. \tag{7.17}$$

If the base of $\log$ is between 0 and 1, then inequality signs in (7.17) are reversed.

In all these inequalities $i+j$ means $i+j-n$ in case of $i+j>n$.

7.3 Cyclic improvemnts of inequalities for entropy of Zipf-Mandelbrot law via Hermite interpolating polynomial

In order to give our main results, we consider the following hypotheses for next sections. (M_1) Let $I \subset \mathbb{R}$ be an interval, $\mathbf{x} := (x_1, \ldots, x_n) \in I^n$ and let $p_1, \ldots, p_n$ and $\lambda_1, \ldots, \lambda_k$ represent positive probability distributions for $2 \leq k \leq n$.

(M_2) Let $f : I \to \mathbb{R}$ be a convex function.

Remark 7.3 Under the conditions (M_1), we define

$$J_1(f) = J_1(\mathbf{x}, \mathbf{p}, \lambda; f) := \sum_{i=1}^{n} p_i f(x_i) - C_{dis}(f, \mathbf{x}, \mathbf{p}, \lambda)$$

$$J_2(f) = J_1(\mathbf{x}, \mathbf{p}, \lambda; f) := C_{dis}(f, \mathbf{x}, \mathbf{p}, \lambda) - f\left(\sum_{i=1}^{n} p_i x_i\right)$$

where $f : I \to \mathbb{R}$ is a function. The functionals $f \to J_u(f)$ are linear, $u = 1, 2$, and Theorem 7.1 imply that

$$J_u(f) \geq 0, \quad u = 1, 2$$

if $f : I \to \mathbb{R}$ is a convex function.

Assume (H_1) and $(H_3$-$H_5)$. Then we have the following additional linear functionals

$$J_3(f) = J_3(f, g, \mu, \mathbf{p}, \lambda) := \int_X f \circ g d\mu - C_{int}(f, g, \mu, \mathbf{p}, \lambda) \geq 0,$$

$$J_4(f) = J_4(t, f, g, \mu, \mathbf{p}, \lambda) := \int_X f \circ g d\mu - C_{par}(t, f, g, \mu, \mathbf{p}, \lambda) \geq 0; \ t \in [0, 1],$$

$$J_5(f) = J_5(t, f, g, \mu, \mathbf{p}, \lambda) := C_{int}(f, g, \mu, \mathbf{p}, \lambda) - C_{par}(t, f, g, \mu, \mathbf{p}, \lambda) \geq 0; \quad t \in [0, 1],$$

$$J_6(f) = J_6(t, f, g, \mu, \mathbf{p}, \lambda) := C_{par}(t, f, g, \mu, \mathbf{p}, \lambda) - f\left(\int_X g d\mu\right) \geq 0; \quad t \in [0, 1].$$

For $v = 1, \ldots, 5$, consider the Green functions $G_v : [\alpha_1, \alpha_2] \times [\alpha_1, \alpha_2] \to \mathbb{R}$ defined as

$$G_1(z, r) = \begin{cases} \frac{(\alpha_2 - z)(\alpha_1 - r)}{\alpha_2 - \alpha_1}, & \alpha_1 \leq r \leq z; \\ \frac{(\alpha_2 - r)(\alpha_1 - z)}{\alpha_2 - \alpha_1}, & z \leq r \leq \alpha_2. \end{cases} \tag{7.18}$$

$$G_2(z, r) = \begin{cases} \alpha_1 - r, & \alpha_1 \leq r \leq z, \\ \alpha_1 - z, & z \leq r \leq \alpha_2. \end{cases} \tag{7.19}$$

$$G_3(z,r) = \begin{cases} z-\alpha_2, & \alpha_1 \le r \le z, \\ r-\alpha_2, & z \le r \le \alpha_2. \end{cases} \tag{7.20}$$

$$G_4(z,r) = \begin{cases} z-\alpha_1, & \alpha_1 \le r \le z, \\ r-\alpha_1, & z \le r \le \alpha_2. \end{cases} \tag{7.21}$$

$$G_5(z,r) = \begin{cases} \alpha_2-r, & \alpha_1 \le r \le z, \\ \alpha_2-z, & z \le r \le \alpha_2, \end{cases} \tag{7.22}$$

All these functions are convex and continuous w.r.t both z and r (see [33]).

Remark 7.4 The Green's function $G_1(\cdot,\cdot)$ is called Lagrange Green's function (see [34]). The new Green functions $G_v(\cdot,\cdot)$, $(v=2,3,4,5)$, introduced by Pečarić et al. in [33].

For $I=[\alpha_1,\alpha_2]$, consider the following assumptions .

(A_1) For the linear functionals $J_u(\cdot)$ $(u=1,2)$, assume that $\dfrac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}z_{i+j}}{\sum\limits_{v=0}^{k-1}\lambda_{j+1}p_{i+j}} \in [\alpha_1,\alpha_2]$ for $i=1,\dots m$.

(A_2) For the linear functionals $J_u(\cdot)$ $(u=3,\dots,6)$, assume that $\dfrac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}f\big(z_{i+j}\big)}{\sum\limits_{j=0}^{k-1}\lambda_{v+1}p_{i+j}} \in [\alpha_1,\alpha_2]$ for $i=1,\dots m$.

7.3.1 Extensions of cyclic refinements of Jensen's inequality via Hermite interpolating polynomial

The proof of the results of this section are given in [16]. We start this section by considering the discrete as well as continuous version of cyclic refinements of Jensen's inequality and construct the generalized new identities having real weights utilizing Hermite interpolating polynomial.

Theorem 7.6 *Let $m,k\in\mathbb{N}$, $p_1,\dots,p_m$ and $\lambda_1,\dots,\lambda_k$ be real tuples for $2\le k\le m$, such that $\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\neq 0$ for $i=1,\dots m$ with $\sum\limits_{i=1}^{m}p_i=1$ and $\sum\limits_{j=1}^{k}\lambda_j=1$. Also let $z\in[\alpha_1,\alpha_2]\subset\mathbb{R}$ and $\mathbf{z}\in[\alpha_1,\alpha_2]^m$. Assume $f\in C^n[\alpha_1,\alpha_2]$ and consider interval with points $-\infty<\alpha_1=b_1<b_2\cdots<b_t=\alpha_2<\infty$, $(t\ge 2)$ such that $f(\alpha_1)=f(\alpha_2)$, $f'(\alpha_1)=0=f'(\alpha_2)$ and G_v, $(v=1,\dots,5)$ be the Green functions defined in* (10.4)–(7.22)*, respectively. Then for $u=1,\dots,6$ along with assumptions (A_1) and (A_2), we have the following generalized identities:*

(a)

$$J_u(f(z)) = \sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega} f^{(\sigma)}(b_\omega)J_u\Big(H_{\sigma\omega}(z)\Big) + \int_{\alpha_1}^{\alpha_2} J_u\Big(G_{H,n}(z,r)\Big)f^{(n)}(r)dr. \tag{7.23}$$

(b)

$$J_u(f(z)) = \int_{\alpha_1}^{\alpha_2} J_u\left(G_v(z,r)\right) \sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} f^{(\sigma+2)}(b_\omega) H_{\sigma\omega}(r) dr$$
$$+ \int_{\alpha_1}^{\alpha_2} \int_{\alpha_1}^{\alpha_2} J_u\left(G_v(z,r)\right) G_{H,n-2}(r,\xi)) f^{(n)}(\xi) d\xi dr \quad (7.24)$$

where $H_{\sigma\omega}$ are Hermite basis and $G_{H,n}(z,r)$ be the Hermite Green function (see [32]).

Now we obtain extensions and improvements of discrete and integral cyclic Jensen type linear functionals, with real weights.

Theorem 7.7 *Consider f be n-convex function along with the suppositions of Theorem 7.6. Then we conclude the following results:*

(a) If for all $u = 1,\dots,6$,

$$J_u\left(G_{H,n}(z,r)\right) \geq 0, \quad r \in [\alpha_1, \alpha_2] \tag{7.25}$$

holds, then we have

$$J_u(f(z)) \geq \sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} f^{(\sigma)}(b_\omega) J_u\left(H_{\sigma\omega}(z)\right) \tag{7.26}$$

for $u = 1,\dots,6$.

(b) If for all $u = 1,\dots,6$ and $v = 1,\dots,5$

$$J_u\left(G_v(z,r)\right) \geq 0, \ r \in [\alpha_1, \alpha_2] \tag{7.27}$$

holds, provided that s_ω is odd for each $\omega = 2,3,4,\cdots,t$, then

$$J_u(f(z)) \geq \int_{\alpha_1}^{\alpha_2} J_u\left(G_v(z,r)\right) \sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} f^{(\sigma+2)}(b_\omega) H_{\sigma\omega}(r) dr. \tag{7.28}$$

for $u = 1,\dots,6$.

(c) If (7.27) holds for all $u = 1,\dots,6$ and $v = 1,\dots,5$, provided that s_ω is odd for each $\omega = 2,3,4,\cdots,t-1$ and s_t is even then (7.28) holds in reverse direction for $u = 1,\dots,6$.

We will finish the present section by the following generalizations of cyclic refinements of Jensen inequalities:

Theorem 7.8 *If the assumptions of Theorem 7.6 be fulfilled with additional conditions that $p_1,\ldots,p_m$ and $\lambda_1,\ldots,\lambda_k$ be non negative tuples for $2 \le k \le m$, such that $\sum_{i=1}^{m} p_i = 1$ and $\sum_{j=1}^{k} \lambda_j = 1$. Then for $\psi : [\alpha_1, \alpha_2] \to \mathbb{R}$ being n-convex function, we conclude the following results:*

(a) If (7.26) is valid along with the function

$$\Gamma(z) := \sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} H_{\sigma\omega}(z) f^{(\sigma)}(b_\omega). \tag{7.29}$$

to be convex, the right side of (7.26) is non negative, means

$$J_u(\psi) \ge 0, \qquad u = 1,\ldots,6. \tag{7.30}$$

(b) If s_ω to be odd for each $\omega = 2,3,4,\cdots,t$, (7.28) holds. Further

$$\sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} H_{\sigma\omega}(r) f^{(\sigma+2)}(b_\omega) \ge 0. \tag{7.31}$$

the right side of (7.28) is non negative, particularly (7.30) is establish for all $u = 1,\ldots,6$ and $v = 1,\ldots,5$..

(c) Inequality (7.28) holds reversely if s_ω is odd for each $\omega = 2,3,4,\cdots,t-1$ and s_t is even. Moreover, let (7.31) holds in reverse direction then reverse of (7.30) holds for all $u = 1,\ldots,6$ and $v = 1,\ldots,5$.

7.3.2 Cyclic improvements of inequalities for entropy of Zipf-Mandelbrot law via Hermite polynomial

Remark 7.5 Now as a consequences of Theorem 7.7 we consider the discrete extensions of cyclic refinements of Jensen's inequalities for $(u = 1)$, from (7.26) with respect to n-convex function f in the explicit form:

$$\sum_{i=1}^{m} p_i f(z_i) - \sum_{i=1}^{m} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) f \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} z_{i+j}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) \ge$$
$$\left(\sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} f^{(\sigma)}(b_\omega) \right) \times$$
$$\left(\sum_{i=1}^{m} p_i H_{\sigma\omega}(z_i) - \sum_{i=1}^{m} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) H_{\sigma\omega} \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} z_{i+j}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) \right), \tag{7.32}$$

where $H_{\sigma\omega}$ are Hermite basis.

Theorem 7.9 *Let $m,k \in \mathbb{N}$ $(2 \le k \le m)$, $\lambda_1,\ldots,\lambda_k$ be positive probability distributions. Let $\mathbf{p} := (p_1,\ldots,p_m) \in \mathbb{R}^m$, and $\mathbf{q} := (q_1,\ldots,q_m) \in (0,\infty)^m$ such that*

$$\frac{p_i}{q_i} \in [\alpha_1,\alpha_2], \quad i = 1,\ldots,m.$$

Also let $f \in C^n[\alpha_1,\alpha_2]$ and consider interval with points $-\infty < \alpha_1 = b_1 < b_2 \cdots < b_t = \alpha_2 < \infty$, $(t \ge 2)$ such that f is n-convex function. Then the following inequalities hold:

$$\hat{I}_f(\mathbf{p},\mathbf{q}) \ge \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) f\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right) + \left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega} f^{(\sigma)}(b_\omega)\right) \times$$
$$\left(\sum_{i=1}^{m} q_i H_{\sigma\omega}\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) H_{\sigma\omega}\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)\right). \tag{7.33}$$

Proof. Replacing p_i with q_i and z_i with $\frac{p_i}{q_i}$ for $(i = 1,\ldots,m)$ in (7.32) , we get (7.33). □

We now explore two exceptional cases of the previous result.
One corresponds to the entropy of a discrete probability distribution.

Corollary 7.7 *Let $m,k \in \mathbb{N}$ $(2 \le k \le m)$, $\lambda_1,\ldots,\lambda_k$ be positive probability distributions.*

(a) If $\mathbf{q} := (q_1,\ldots,q_m) \in (0,\infty)^m$ and $(n = even)$, then

$$\sum_{i=1}^{m} q_i \ln q_i \ge \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)\ln\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) + \left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right) \times$$
$$\left(\sum_{i=1}^{m} q_i H_{\sigma\omega}\left(\frac{1}{q_i}\right) - \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) H_{\sigma\omega}\left(\frac{1}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)\right). \tag{7.34}$$

(b) If $\mathbf{q} := (q_1,\ldots,q_m)$ is a positive probability distribution and $(n = even)$, then we get the bounds for the Shannon entropy of $\mathbf{q}$.

$$H(\mathbf{q}) \leq -\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)\ln\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) - \left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right)\times$$
$$\left(\sum_{i=1}^{m}q_i H_{\sigma\omega}\left(\frac{1}{q_i}\right) - \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)H_{\sigma\omega}\left(\frac{1}{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)\right). \quad (7.35)$$

If $(n = odd)$, then (7.34) *and* (7.35) *hold in reverse directions.*

Proof.

(a) Using $f(x) := -\ln x$ and $\mathbf{p} := (1,1,\ldots,1)$ in Theorem 7.9, we get the required results.

(b) It is a specific case of (a).

□

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

Corollary 7.8 *Let $m,k \in \mathbb{N}$ $(2 \leq k \leq m)$, $\lambda_1,\ldots,\lambda_k$ be positive probability distributions.*

(a) If $\mathbf{q} := (q_1,\ldots,q_m), \mathbf{p} := (p_1,\ldots,p_m) \in (0,\infty)^m$ and $(n = even)$, then

$$\sum_{i=1}^{m}q_i\ln\left(\frac{q_i}{p_i}\right) \geq \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)\ln\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) + \left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right)\times$$
$$\left(\sum_{i=1}^{m}q_i H_{\sigma\omega}\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)H_{\sigma\omega}\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)\right). \quad (7.36)$$

(b) If If $\mathbf{q} := (q_1,\ldots,q_m), \mathbf{p} := (p_1,\ldots,p_m)$ are positive probability distributions and $(n = even)$, then we have

$$D(\mathbf{q}\parallel\mathbf{p}) \geq \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)\ln\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) +$$

$$\left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right)\times$$
$$\left(\sum_{i=1}^{m} q_i H_{\sigma\omega}\left(\frac{p_i}{q_i}\right)-\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)H_{\sigma\omega}\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)\right). \tag{7.37}$$

If $(n = odd)$, then (7.36) *and* (7.37) *hold in reverse directions.*

Proof.

(a) Using $f(x) := -\ln x$ in Theorem 7.9, we get the desired results.

(b) It is particular case of (a).

□

Let $m \in \{1,2,\ldots\}$, $t \geq 0$, $s > 0$, then **Zipf-Mandelbrot entropy** can be given as:

$$Z(H,t,s) = \frac{s}{H_{m,t,s}}\sum_{i=1}^{m}\frac{\ln(i+t)}{(i+t)^s} + \ln(H_{m,t,s}). \tag{7.38}$$

Consider

$$q_i = f(i;m,t,s) = \frac{1}{((i+t)^s H_{m,t,s})}. \tag{7.39}$$

Now we state our results involving entropy introduced by Mandelbrot Law:

Theorem 7.10 *Let $m,k \in \mathbb{N}$ $(2 \leq k \leq m)$, $\lambda_1,\ldots,\lambda_k$ be positive probability distributions and $\mathbf{q}$ be as defined in* (7.39) *by Zipf-Mandelbrot law with parameters $m \in \{1,2,\ldots\}$, $c \geq 0$, $d > 0$. For $(n = even)$, the following holds*

$$\begin{aligned}
H(\mathbf{q}) &= Z(H,c,d)\\
&\leq -\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\frac{\lambda_{j+1}}{((i+j+c)^d H_{m,c,d})}\right)\ln\left(\frac{1}{H_{m,c,d}}\sum_{j=0}^{k-1}\frac{\lambda_{j+1}}{((i+j+t)^s)}\right)-\\
&\left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right)\left(\sum_{i=1}^{m}\frac{1}{((i+c)^d H_{m,c,d})}H_{\sigma\omega}\left(((i+c)^d H_{m,c,d})\right)\right)+\\
&\left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right)\left(\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\frac{\lambda_{j+1}}{((i+j+c)^d H_{m,c,d})}\right)H_{\sigma\omega}\left(\frac{1}{\sum_{j=0}^{k-1}\frac{\lambda_{j+1}}{((i+j+c)^d H_{m,c,d})}}\right)\right).
\end{aligned} \tag{7.40}$$

If $(n = odd)$, then (7.40) *holds in reverse direction.*

Proof. Substituting this $q_i = \frac{1}{((i+c)^d H_{m,c,d})}$ in Corollary 7.7(b), we get the desired result. Since it is interesting to see that $\sum\limits_{i=1}^{m} q_i = 1$. Moreover using above q_i in Shannon entropy (7.3), we get Mandelbrot entropy(7.38). □

Corollary 7.9 *Let $m,k \in \mathbb{N}$ $(2 \le k \le m)$, $\lambda_1,\ldots,\lambda_k$ be positive probability distributions and for $c_1, c_2 \in [0,\infty)$, $d_1, d_2 > 0$, let $H_{m,c_1,d_1} = \frac{1}{(i+c_1)^{d_1}}$ and $H_{m,c_2,d_2} = \frac{1}{(i+c_2)^{d_2}}$. Now using $q_i = \dfrac{1}{(i+c_1)^{d_1} H_{m,c_1,d_1}}$ and $p_i = \dfrac{1}{(i+c_2)^{d_2} H_{m,c_2,d_2}}$ in Corollary 7.8(b), with $(n = even)$, then the following holds*

$$
\begin{aligned}
D(\mathbf{q} \parallel \mathbf{p}) &= \sum_{i=1}^{m} \frac{1}{(i+c_1)^{d_1} H_{m,c_1,d_1}} \ln\left(\frac{(i+c_2)^{d_2} H_{m,c_2,d_2}}{(i+c_1)^{d_1} H_{m,c_1,d_1}}\right) \\
&\ge \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+j+c_1)^{d_1} H_{m,c_1,d_1}}\right) \ln\left(\frac{\sum\limits_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+c_1)^{d_1} H_{m,c_1,d_1}}}{\sum\limits_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+c_2)^{d_2} H_{m,c_2,d_2}}}\right) \\
&+ \left(\sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} \frac{(-1)^\sigma (\sigma-1)!}{(b_\omega)^\sigma}\right)\left(\sum_{i=1}^{m} \frac{1}{((i+c_1)^{d_1} H_{m,c_1,d_1})} H_{\sigma\omega}\left(\frac{((i+c_2)_2^d H_{m,c_2,d_2})}{((i+c_1)^d H_{m,c_1,d_1})}\right)\right) \\
&- \left(\sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} \frac{(-1)^\sigma (\sigma-1)!}{(b_\omega)^\sigma}\right) \times \\
&\left(\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+j+c_1)^{d_1} H_{m,c_1,d_1}}\right) H_{\sigma\omega}\left(\frac{\sum\limits_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+c_1)^{d_1} H_{m,c_1,d_1}}}{\sum\limits_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+c_2)^{d_2} H_{m,c_2,d_2}}}\right)\right)
\end{aligned}
\tag{7.41}
$$

If $(n = odd)$, then (7.41) *holds in reverse direction.*

Remark 7.6 It is interesting to note that, in the similar passion we are able to construct different estimations of f-divergences along with their applications to Shannon and Mandelbrot entropies using the other inequalities for n-convex functions constructed in Theorem 7.7 for discrete case of cyclic refinements of Jensen inequality.

Remark 7.7 We left for reader interest to construct upper bounds for Shannon, Relative and Mandelbrot entropies by considering **Type**$(\eta, n-\eta)$**C** and **Two-point TC** instead of **HC** in the above results.

7.4 A refinement and an exact equality condition for the basic inequality of f-divergences

Measures of dissimilarity between probability measures play important role in probability theory, especially in information theory and in mathematical statistics. Many divergence measures for this purpose have been introduced and studied (see for example Vajda [14]). Among them f-divergences were introduced by Csiszár [2] and [37] and independently by Ali and Silvey [1]. Remarkable divergences can be found among f-divergences, such as the information divergence, the Pearson or χ^2-divergence, the Hellinger distance and total variational distance. There are a lot of papers dealing with f-divergence inequalities (see Dragomir [39], Dembo, Cover, and Thomas [4] and Sason and Verdú [50]). These inequalities are very useful and applicable in information theory.

One of the basic inequalities is (see Liese and Vajda [45])

$$D_f(P,Q) \geq f(1).$$

In this section we give a refinement and a precise equality condition for this inequality. Some applications for discrete distributions, for the Shannon entropy, and some examples are given.

7.4.1 Construction of the equality conditions and related results of classical integral Jensen's inequality

The classical Jensen's inequality is well known (see [7]).

Theorem 7.11 *Let g be an integrable function on a probability space $(Y, \mathscr{B}, \nu)$ taking values in an interval $I \subset \mathbb{R}$. Then $\int_Y g d\nu$ lies in I. If f is a convex function on I such that $f \circ g$ is ν-integrable, then*

$$f\left(\int_Y g d\nu\right) \leq \int_Y f \circ g d\nu. \tag{7.42}$$

The following approach to give a necessary and sufficient condition for equality in this inequality may be new. First, we introduce the next definition.

Definition 7.7 *Let $(Y, \mathscr{B}, \nu)$ be a probability space, and let g be a real measurable function defined almost everywhere on Y. We denote by $essint_\nu(g)$ the smallest interval in $\mathbb{R}$ for which*

$$\nu(g \in essint_\nu(g)) = 1.$$

Remark 7.8 (a) Obviously, the endpoints of $\operatorname{essint}_\nu(g)$ are the essential infimum ($\operatorname{essinf}_\nu(g)$) and the essential supremum of g, and either of them belong to $\operatorname{essint}_\nu(g)$ exactly if g takes this value with positive probability.

(b) It is easy to see that either $\operatorname{essint}_\nu(g) = \left\{ \int_Y g d\nu \right\}$ (in this case g is constant ν-a.e.) or $\int_Y g d\nu$ is an inner point of $\operatorname{essint}_\nu(g)$.

(c) The interval $\operatorname{essinf}_\nu(g)$ is connected with the essential range of g, but not the same set (for example, the essential range of g is always closed, and not an interval in general).

Lemma 7.1 *Assume the conditions of Theorem 7.11 are satisfied. Equality holds in (7.42) if and only if f is affine on $\operatorname{essint}_\nu(g)$.*

Proof. It is easy to see that the condition is sufficient for equality in (7.42).

Conversely, if $\operatorname{essint}_\nu(g)$ contains only one point, then it is trivial, so we can assume that $m := \int_Y g d\nu$ is an inner point of $\operatorname{essint}_\nu(g)$. Let

$$l : \mathbb{R} \to \mathbb{R}, \quad l(t) = f'_+(m)(t-m) + f(m).$$

If f is not affine on $\operatorname{essint}_\nu(g)$, then by the convexity of f, there is a point $t_1 \in \operatorname{essint}_\nu(g)$ such that $f(t_1) > l(t_1)$. Suppose $t_1 > m$ (the case $t_1 < m$ can be handled similarly). Since f is convex, $f(t) \geq l(t)$ $(t \in I)$ and $f(t) > l(t)$ $(t \in I,\ t \geq t_1)$. It follows by using $\nu(g > t_1) > 0$, that

$$\int_Y f \circ g d\nu = \int_{(g<t_1)} f \circ g d\nu + \int_{(g \geq t_1)} f \circ g d\nu$$

$$\geq \int_{(g<t_1)} l \circ g d\nu + \int_{(g \geq t_1)} f \circ g d\nu > \int_Y l \circ g d\nu = f(m),$$

which is a contradiction.

The proof is complete. □

The next refinement of the Jensen's inequality can be found in Horváth [8].

Theorem 7.12 *Let $I \subset \mathbb{R}$ be an interval, and let $f : I \to \mathbb{R}$ be a convex function. Let $(Y, \mathscr{B}, \nu)$ be a probability space, and let $g : Y \to I$ be a ν-integrable function such that $f \circ g$ is also ν-integrable. Suppose that $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n} \alpha_i = 1$. Then*

(a)

$$f\left(\int_Y g d\nu\right) \leq \int_{Y^n} f\left(\sum_{i=1}^{n} \alpha_i g(x_i)\right) d\nu^n(x_1, \ldots, x_n) \leq \int_Y f \circ g d\nu.$$

(b)

$$\int_{Y^{n+1}} f\left(\frac{1}{n+1}\sum_{i=1}^{n+1} g(x_i)\right) d\nu^{n+1}(x_1,\ldots,x_{n+1})$$

$$\leq \int_{Y^n} f\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right) d\nu^{n}(x_1,\ldots,x_n) \leq \int_{Y^n} f\left(\sum_{i=1}^{n} \alpha_i g(x_i)\right) d\nu^{n}(x_1,\ldots,x_n).$$

By analyzing the proof of the previous result, it can be seen that the hypothesis "$f \circ g$ is ν-integrable" can be weaken.

Theorem 7.13 *Let $I \subset \mathbb{R}$ be an interval, and let $f : I \to \mathbb{R}$ be a convex function. Let $(Y, \mathscr{B}, \nu)$ be a probability space, and let $g : Y \to I$ be a ν-integrable function such that the integral $\int_Y f \circ g d\nu$ exists in $]-\infty, \infty]$. Suppose that $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n} \alpha_i = 1$. Then the assertions of Theorem 7.12 remain true.*

We assume throughout that the probability measures P and Q are defined on a fixed measurable space $(X, \mathscr{A})$. It is also assumed that P and Q are absolutely continuous with respect to a σ-finite measure μ on $\mathscr{A}$. The densities (or Radon-Nikodym derivatives) of P and Q with respect to μ are denoted by p and q, respectively. These densities are μ-almost everywhere uniquely determined.

Let

$$F := \{f :]0, \infty[\to \mathbb{R} \mid f \text{ is convex}\},$$

and define for every $f \in F$ the function

$$f^* :]0, \infty[\to \mathbb{R}, \quad f^*(t) := t f\left(\frac{1}{t}\right).$$

If $f \in F$, then either f is monotonic or there exists a point $t_0 \in]0, \infty[$ such that f is decreasing on $]0, t_0[$. This implies that the limit

$$\lim_{t \to 0+} f(t)$$

exists in $]-\infty, \infty]$, and

$$f(0) := \lim_{t \to 0+} f(t)$$

extends f into a convex function on $[0, \infty[$. The extended function is continuous and has finite left and right derivatives at each point of $]0, \infty[$.

It is well known that for every $f \in F$ the function f^* also belongs to F, and therefore

$$f^*(0) := \lim_{t \to 0+} f^*(t) = \lim_{u \to \infty} \frac{f(u)}{u}.$$

We need the following simple property of functions belonging to F.

Lemma 7.2 *If $f \in F$, then $f^*(0) \geq f'_+(1)$. This inequality becomes an equality if and only if*

$$f(t) = f'_+(1)(t-1) + f(1), \quad t \geq 1. \tag{7.43}$$

Proof. Since f is convex,

$$f(t) \geq f'_+(1)(t-1) + f(1), \quad t \geq 1,$$

and therefore

$$f^*(0) = \lim_{t\to\infty} \frac{f(t)}{t} \geq f'_+(1).$$

If (7.43) is satisfied, then obviously $f^*(0) = f'_+(1)$.

If there exists $t_1 > 1$ such that $f'_+(t_1) > f'_+(1)$, then by the convexity of f,

$$f(t) \geq f'_+(t_1)(t-t_1) + f(t_1), \quad t \geq t_1,$$

and hence $f^*(0) > f'_+(1)$. It follows that $f^*(0) = f'_+(1)$ implies

$$f'_+(t) = f'_+(1), \quad t \geq t_1,$$

and this gives (7.43) (see [43] 1.6.2 Corollary 2).

The proof is complete. □

The next result prepares the notion of f-divergence of probability measures.

Lemma 7.3 *For every $f \in F$ the integral*

$$\int_{(q>0)} q(\omega) f\left(\frac{p(\omega)}{q(\omega)}\right) d\mu(\omega)$$

exists and it belongs to the interval $]-\infty, \infty]$.

Proof. Since f is convex,

$$f(t) \geq f'_+(1)(t-1) + f(1), \quad t \geq 0.$$

This implies that for all $\omega \in (q>0)$

$$q(\omega) f\left(\frac{p(\omega)}{q(\omega)}\right) \geq h(\omega) := f'_+(1)(p(\omega) - q(\omega)) + f(1) q(\omega). \tag{7.44}$$

Elementary considerations show that the function h is μ-integrable over $(q>0)$, and this gives the result by (7.44).

The proof is complete. □

Now we introduce the notion of f-divergence.

Definition 7.8 *For every $f \in F$ we define the f-divergence of P and Q by*

$$D_f(P,Q) := \int_X q(\omega) f\left(\frac{p(\omega)}{q(\omega)}\right) d\mu(\omega),$$

where the following conventions are used

$$0f\left(\frac{x}{0}\right) := xf^*(0) \text{ if } x > 0, \quad 0f\left(\frac{0}{0}\right) = 0f^*(0) := 0. \tag{7.45}$$

Remark 7.9 (a) For every $f \in F$ the perspective $\hat{f} :]0,\infty[\times]0,\infty[\to \mathbb{R}$ of f is defined by

$$\hat{f}(x,y) := yf\left(\frac{x}{y}\right).$$

Then (see [49]) $\hat{f}$ is also a convex function. Vajda [14] proved that (7.45) is the unique rule leading to convex and lower semicontinuous extension of $\hat{f}$ to the set

$$\left\{(x,y) \in \mathbb{R}^2 \mid x,y \geq 0\right\}.$$

(b) Since $f^*(0) \in]-\infty,\infty]$, Lemma 7.3 shows that $D_f(P,Q)$ exists in $]-\infty,\infty]$ and

$$D_f(P,Q) = \int_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega) + f^*(0) P(q=0). \tag{7.46}$$

It follows that if P is absolutely continuous with respect to Q, then

$$D_f(P,Q) = \int_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega).$$

Various divergences in information theory and statistics are special cases of the f-divergence. We illustrate this by some examples.

(a) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(t) = t\ln(t)$ in (7.46), the information divergence is obtained

$$I(P,Q) = \int_{(q>0)} p(\omega) \ln\left(\frac{p(\omega)}{q(\omega)}\right) d\mu(\omega) + \infty P(q=0). \tag{7.47}$$

(b) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(t) = (t-1)^2$ in (7.46), the Pearson or χ^2-divergence is obtained

$$\chi^2(P,Q) = \int_{(q>0)} \frac{(p(\omega) - q(\omega))^2}{q(\omega)} d\mu(\omega) + \infty P(q=0). \tag{7.48}$$

(c) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(t) = \left(\sqrt{t} - 1\right)^2$ in (7.46), the Hellinger distance is obtained

$$H^2(P,Q) = \int_X \left(\sqrt{p(\omega)} - \sqrt{q(\omega)}\right)^2 d\mu(\omega). \tag{7.49}$$

(d) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(t) = |t-1|$ in (7.46), the total variational distance is obtained

$$V(P,Q) = \int_X |p(\omega) - q(\omega)| \mu(\omega). \tag{7.50}$$

We need the following lemma.

Lemma 7.4 *Let $t_0 := P(q > 0)$.*

(a) For every $\varepsilon > 0$

$$Q\left(\frac{p}{q} < t_0 + \varepsilon,\ q > 0\right) > 0.$$

(b)

$$essinf_Q\left(\frac{p}{q}\right) \leq t_0$$

Proof. (a) Obviously,

$$Q\left(\frac{p}{q} < t_0 + \varepsilon,\ q > 0\right) = 1 - Q\left(\frac{p}{q} \geq t_0 + \varepsilon,\ q > 0\right).$$

The result follows from this, since

$$Q\left(\frac{p}{q} \geq t_0 + \varepsilon,\ q > 0\right) = \int_X q 1_{\left(\frac{p}{q} \geq t_0+\varepsilon,\ q>0\right)} d\mu \leq \int_{(q>0)} \frac{1}{t_0+\varepsilon} p d\mu$$

$$= \frac{t_0}{t_0+\varepsilon} < 1.$$

(b) It comes from (a).

The proof is complete. □

The following result contains a key property of f-divergences. We give a simple proof which emphasizes the importance of the convexity of f, and give an exact equality condition.

Theorem 7.14 *(a) For every $f \in F$*

$$D_f(P,Q) \geq f(1). \tag{7.51}$$

(b) Assume $P(q = 0) = 0$. Then equality holds in (7.51) if and only if f is affine on $essint_Q\left(\frac{p}{q}\right)$.

(c) Assume $P(q = 0) > 0$. Then equality holds in (7.51) if and only if f is affine on $essint_Q\left(\frac{p}{q}\right) \cup [1,\infty[$.

Proof. (a) If $D_f(P,Q) = \infty$, then (7.51) is obvious.

If $D_f(P,Q) \in \mathbb{R}$, then the integral

$$\int_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega) \tag{7.52}$$

is finite, and therefore either $Q(p=0) = 0$ or $Q(p=0) > 0$ and $f(0)$ is finite. It follows that Jensen's inequality can be applied to this integral, and we have

$$D_f(P,Q) \geq f\left(\int_{(q>0)} p d\mu\right) + f^*(0)P(q=0) \tag{7.53}$$

$$= f(P(q>0)) + f^*(0)P(q=0). \tag{7.54}$$

Let $t_0 := P(q>0)$. By using Lemma 7.2, $t_0 \in [0,1]$, and the convexity of f, it follows from (7.54) that

$$D_f(P,Q) \geq f(t_0) + f'_+(1)(1-t_0) \tag{7.55}$$

$$\geq f(1) + f'_+(1)(t_0-1) + f'_+(1)(1-t_0) = f(1). \tag{7.56}$$

(b) If $D_f(P,Q) = f(1)$, then $D_f(P,Q)$ is finite.

Assume $P(q=0) = 0$. Then by (7.53) and (7.54), $D_f(P,Q) = f(1)$ is satisfied if and only if equality holds in the Jensen's inequality. Lemma 7.1 shows that this happens exactly if f is affine on $\text{essint}_Q\left(\frac{p}{q}\right)$.

(c) Assume $P(q=0) > 0$. Then (7.53), (7.54), (7.55) and (7.56) yield that there must be equality in the Jensen's inequality, $f^*(0) = f'_+(1)$, and

$$f(t_0) = f(1) + f'_+(1)(t_0-1). \tag{7.57}$$

By Lemma 7.1 and Lemma 7.2, the first two equality conditions are satisfied exactly if f is affine on $\text{essint}_Q\left(\frac{p}{q}\right) \cup [1,\infty[$.

Now assume that f is affine on $\text{essint}_Q\left(\frac{p}{q}\right) \cup [1,\infty[$. In case of $t_0 > 0$, Lemma 7.4 (b) and the continuity of f at t_0 show that (7.57) also holds. In case of $t_0 = 0$, it is easy to see that $Q\left(\frac{p}{q} = 0\right) = 1$, and hence $0 \in \text{essint}_Q\left(\frac{p}{q}\right)$ which implies (7.57) too.

The proof is complete. $\square$

Remark 7.10 (a) Consider the subclass $F_1 \subset F$ such that $f \in F_1$ satisfies $f(1) = 0$. In this case inequality (7.51) has the usual form

$$D_f(P,Q) \geq 0.$$

(b) The usual equality condition is the next (see [45]): if f is strictly convex at 1, then $D_f(P,Q) = f(1)$ holds if and only if $P = Q$. Theorem 7.14 (b) and (c) give more precise conditions.

7.4.2 Refinements of basic inequality in f-divergences and related results

Suppose that $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n} \alpha_i = 1$. Let

$$\mathscr{A}^n := \mathscr{A} \otimes \ldots \otimes \mathscr{A}, \quad \text{with } n \text{ factors},$$

and define the probability measures Q^n and R on $\mathscr{A}^n$ by

$$Q^n := Q \otimes \ldots \otimes Q, \quad \text{with } n \text{ factors},$$

and

$$R_\alpha := \sum_{i=1}^{n} \alpha_i Q \otimes \ldots \otimes Q \otimes \overset{i}{\breve{P}} \otimes Q \otimes \ldots \otimes Q.$$

In case of $\alpha_i = \frac{1}{n}$ $(i = 1, \ldots, n)$ the probability measure R_α will be denoted by R_n.

These measures are absolutely continuous with respect to μ^n on $\mathscr{A}^n$. The densities of R and Q^n with respect to μ^n are

$$\bigotimes_{i=1}^{n} q : X^n \to \mathbb{R}, \quad (\omega_1, \ldots, \omega_n) \to \prod_{i=1}^{n} q(\omega_i),$$

and

$$(\omega_1, \ldots, \omega_n) \to \sum_{i=1}^{n} \alpha_i q(\omega_1) \ldots \overset{i}{\breve{p}}(\omega_i) \ldots q(\omega_n), \quad (\omega_1, \ldots, \omega_n) \in X^n,$$

respectively.

It is easy to calculate that

$$R_\alpha \left(\bigotimes_{i=1}^{n} q = 0 \right) = 1 - R_\alpha \left(\bigotimes_{i=1}^{n} q > 0 \right) = 1 - R_\alpha \left((q > 0)^n \right)$$

$$= 1 - \sum_{i=1}^{n} \alpha_i Q(q > 0)^{n-1} P(q > 0) = 1 - P(q > 0) = P(q = 0).$$

It follows that for every $f \in F$

$$D_f(R_\alpha, Q^n) = \int_{(q>0)^n} f \left(\frac{\sum_{i=1}^{n} \alpha_i q(\omega_1) \ldots p(\omega_i) \ldots q(\omega_n)}{\prod_{i=1}^{n} q(\omega_i)} \right) dQ^n(\omega_1, \ldots, \omega_n)$$

$$+ f^*(0) R_\alpha \left(\bigotimes_{i=1}^{n} q = 0 \right)$$

$$= \int_{(q>0)^n} f\left(\sum_{i=1}^{n}\alpha_i \frac{p(\omega_i)}{q(\omega_i)}\right) dQ^n(\omega_1,\dots,\omega_n) + f^*(0)P(q=0) \tag{7.58}$$

$$= \int_{(q>0)^n} \prod_{i=1}^{n} q(\omega_i) f\left(\sum_{i=1}^{n}\alpha_i \frac{p(\omega_i)}{q(\omega_i)}\right) d\mu^n(\omega_1,\dots,\omega_n) + f^*(0)P(q=0).$$

By applying Theorem 7.12, we obtain some refinements of the basic inequality 7.51.

Theorem 7.15 *Suppose that $\alpha_1,\dots,\alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n}\alpha_i = 1$. If $f \in F$, then*

(a)

$$D_f(P,Q) \geq D_f(R_\alpha, Q^n) \geq D_f(R_n, Q^n) \geq f(1). \tag{7.59}$$

(b)

$$D_f(P,Q) = D_f\left(R_1, Q^1\right)$$

$$\geq \ldots \geq D_f(R_m, Q^m) \geq D_f\left(R_{m+1}, Q^{m+1}\right) \geq \ldots \geq f(1), \quad m \geq 1.$$

Proof. (a) The third inequality in (7.59) comes from Theorem 7.14.

So it remains to prove the first two inequalities in (7.59). By (7.46) and (7.58), it is enough to show that

$$\int_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega) \geq \int_{(q>0)^n} f\left(\sum_{i=1}^{n}\alpha_i \frac{p(\omega_i)}{q(\omega_i)}\right) dQ^n(\omega_1,\dots,\omega_n) \tag{7.60}$$

$$\geq \int_{(q>0)^n} f\left(\frac{1}{n}\sum_{i=1}^{n}\frac{p(\omega_i)}{q(\omega_i)}\right) dQ^n(\omega_1,\dots,\omega_n),$$

which is an immediate consequence of Theorem 7.13.

(b) We can proceed similarly as in (a).

The proof is complete. □

By considering the special f-divergences (7.47-7.50), we have after each other

(a) the information divergence

$$I(R_\alpha, Q^n) = \infty P(q=0)$$

$$+ \int_{(q>0)^n} \sum_{i=1}^{n} \left(\alpha_i p(\omega_i) \prod_{\substack{j=1 \\ j\neq i}}^{n} q(\omega_j)\right) \ln\left(\sum_{i=1}^{n}\alpha_i \frac{p(\omega_i)}{q(\omega_i)}\right) d\mu^n(\omega_1,\dots,\omega_n),$$

(b) the Pearson divergence

$$\chi^2 (R_\alpha, Q^n) =$$

$$= \int\limits_{(q>0)^n} \prod_{i=1}^{n} q(\omega_i) \left(\sum_{i=1}^{n} \alpha_i \frac{p(\omega_i) - q(\omega_i)}{q(\omega_i)} \right)^2 d\mu^n (\omega_1, \ldots, \omega_n) + \infty P(q=0),$$

(c) the Hellinger distance

$$H^2 (R_\alpha, Q^n) = \int\limits_{(q>0)^n} \prod_{i=1}^{n} q(\omega_i) \left(\left(\sum_{i=1}^{n} \alpha_i \frac{p(\omega_i)}{q(\omega_i)} \right)^{1/2} - 1 \right)^2 d\mu^n (\omega_1, \ldots, \omega_n),$$

(d) the total variational distance

$$V (R_\alpha, Q^n) = \int\limits_{(q>0)^n} \prod_{i=1}^{n} q(\omega_i) \left| \sum_{i=1}^{n} \alpha_i \frac{p(\omega_i) - q(\omega_i)}{q(\omega_i)} \right| d\mu^n (\omega_1, \ldots, \omega_n).$$

Now, we consider the special case, important in many applications, in which P and Q are discrete distributions.

Denote T either the set $\{1, \ldots, k\}$ with a fixed positive integer k, or the set $\{1, 2, \ldots\}$. We say that P and Q are derived from the positive probability distributions $p := (p_i)_{i \in T}$ and $q := (q_i)_{i \in T}$, respectively, if p_i, $q_i > 0$ $(i \in T)$, and $\sum\limits_{i \in T} p_i = \sum\limits_{i \in T} q_i = 1$. In this case $X = T$, $\mathscr{A}$ is the power set of T, and μ is the counting measure on $\mathscr{A}$.

Corollary 7.10 *Suppose that $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers with $\sum\limits_{i=1}^{n} \alpha_i = 1$. Suppose also that P and Q are derived from the positive probability distributions $(p_i)_{i \in T}$ and $(q_i)_{i \in T}$, respectively. If $f \in F$, then*

(a)

$$D_f (P, Q) = \sum_{i \in T} q_i f\left(\frac{p_i}{q_i} \right) \geq \sum_{(i_1, \ldots, i_n) \in T^n} \prod_{j=1}^{n} q_{i_j} f\left(\sum_{j=1}^{n} \alpha_j \frac{p_{i_j}}{q_{i_j}} \right)$$

$$\geq \sum_{(i_1, \ldots, i_n) \in T^n} \prod_{j=1}^{n} q_{i_j} f\left(\frac{1}{n} \sum_{j=1}^{n} \frac{p_{i_j}}{q_{i_j}} \right) \geq f(1).$$

(b)

$$D_f (P, Q) \geq \ldots \geq \sum_{(i_1, \ldots, i_n) \in T^n} \prod_{j=1}^{n} q_{i_j} f\left(\frac{1}{n} \sum_{j=1}^{n} \frac{p_{i_j}}{q_{i_j}} \right)$$

$$\geq \sum_{(i_1, \ldots, i_{n+1}) \in T^{n+1}} \prod_{j=1}^{n+1} q_{i_j} f\left(\frac{1}{n+1} \sum_{j=1}^{n+1} \frac{p_{i_j}}{q_{i_j}} \right) \geq \ldots \geq f(1), \quad n \geq 1.$$

Proof. This comes from Theorem 7.15 immediately. □

Finally, we give an example to illustrate the previous result. We consider only Corollary 7.10 (a).

Example 7.1 (a) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(x) = -\ln(x)$ and $p_i = \frac{1}{k}$ $(i = 1,\ldots,k)$ in the previous corollary (in this case $T = \{1,\ldots,k\}$), we have

$$D_f(P,Q) = -\sum_{i=1}^{k} q_i \ln\left(\frac{1}{kq_i}\right) = \ln(k) + \sum_{i=1}^{k} q_i \ln(q_i)$$

$$\geq -\sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\frac{1}{k}\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right) = \ln(k) - \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right)$$

$$\geq -\sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\frac{1}{kn}\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right)$$

$$= \ln(kn) - \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right) \geq 0.$$

It can be obtained from this some refinements of the classical upper estimation for the Shannon entropy

$$H(Q) := -\sum_{i=1}^{k} q_i \ln(q_i) \leq \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right)$$

$$\leq -\ln(n) + \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right) \leq \ln(k).$$

(b) If $f :]0,\infty[\to \mathbb{R}$, $f(x) = x\ln(x)$ in the previous corollary, then we have the following estimations for the information or Kullback–Leibler divergence:

$$I(P,Q) = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right) \geq \sum_{(i_1,\ldots,i_n)\in T^n} \left(\sum_{j=1}^{n} \alpha_j p_{i_j} \prod_{\substack{l=1\\ l\neq j}}^{n} q_{i_l}\right) \ln\left(\sum_{j=1}^{n} \alpha_j \frac{p_{i_j}}{q_{i_j}}\right)$$

$$\geq \frac{1}{n} \sum_{(i_1,\ldots,i_n)\in T^n} \left(\sum_{j=1}^{n} p_{i_j} \prod_{\substack{l=1\\ l\neq j}}^{n} q_{i_l}\right) \ln\left(\frac{1}{n}\sum_{j=1}^{n}\frac{p_{i_j}}{q_{i_j}}\right) \geq 0. \tag{7.61}$$

(c) The Zipf-Mandelbrot law (see Mandelbrot [46] and Zipf [15]) is a discrete probability distribution depends on three parameters $N \in \{1,2,\ldots\}$, $q \in [0,\infty[$ and $s > 0$, and it is defined by

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

where

$$H_{N,q,s} := \sum_{k=1}^{N} \frac{1}{(k+q)^s}.$$

Let P and Q be the Zipf-Mandelbrot law with parameters $N \in \{1,2,\ldots\}$, q_1, $q_2 \in [0,\infty[$ and s_1, $s_2 > 0$, respectively, and let $2 \le k \le N$ be an integer. It follows from the first part of (7.61) with $T = \{1,\ldots,N\}$ that

$$I(P,Q) = \sum_{i=1}^{N} \frac{1}{(i+q_1)^{s_1} H_{N,q_1,s_1}} \log\left(\frac{(i+q_2)^{s_2} H_{N,q_2,s_2}}{(i+q_1)^{s_1} H_{N,q_1,s_1}}\right)$$

$$\ge \sum_{(i_1,\ldots,i_N)\in T^n} \left(\sum_{j=1}^{n} \alpha_j \frac{1}{(i_j+q_1)^{s_1} H_{N,q_1,s_1}} \prod_{\substack{l=1\\ l\neq j}}^{n} \frac{1}{(i_l+q_2)^{s_2} H_{N,q_2,s_2}}\right)$$

$$\times \ln\left(\sum_{j=1}^{n} \alpha_j \frac{(i_j+q_2)^{s_2} H_{N,q_2,s_2}}{(i_j+q_1)^{s_1} H_{N,q_1,s_1}}\right) \ge 0.$$

This is another type of refinement for $I(P,Q)$ than it is given in [9].

Bibliography

[1] M. S. Ali and D. Silvey, *A general class of coefficients of divergence of one distribution from another*, J. Roy. Statist. Soc., ser. B **28** (1966), 131–140.

[2] I. Csiszár, *Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität on Markoffschen Ketten*, Publ. Math. Inst. Hungar. Acad. Sci., ser. A **8** (1963), 84–108.

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

[4] A. Dembo, T.M. Cover and J. A. Thomas, *Information theoretic inequalities*, IEEE Trans. Inf. Theory **37** (1991), 1501–1518.

[5] S. S. Dragomir (Ed.), *Inequalities for Csiszár f-Divergence in Information Theory*, RGMIA Monographs, Victoria University, 2000. (online: http://ajmaa.org/RGMIA/monographs.php/)

[6] T. M. Flett, *Differential Analysis*, Cambridge University Press, 1980.

[7] E. Hewitt and K. R. Stromberg, *Real and Abstract Analysis*, Graduate Text in Mathematics 25, Springer-Verlag, Berlin-Heidelberg-New York, 1965.

[8] L. Horváth, *Inequalities corresponding to the classical Jensen's inequality*, J. Math. Inequal. **3** (2009), 189–200.

[9] L. Horváth, D . Pečarić, J. Pečarić, *Estimations of f- and Rényi divergences by using a cyclic refinement of the Jensen's inequality*, to appear in Bull. Malays. Math. Sci. Soc. 2017. Online publication: doi:10.1007/s40840-017-0526-4

[10] F. Liese and I. Vajda, *On divergences and informations in statistics and information theory*, IEEE Trans. Inf. Theory **52** (2006), 394–4412.

[11] B. Mandelbrot, An informational theory of the statistical structure of language, In Jackson, W., editor, Communication Theory, pp. 486–502. Academic Press, New York (1953)

[12] T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[13] I. Sason and S. Verdú, *f-Divergence Inequalities*, IEEE Trans. Inf. Theory **62** (2016), 5973–6006.

[14] I. Vajda, *Theory of Statistical Inference and Information*, Boston, MA: Kluwer, 1989.

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

[16] N. Mehmood, S. I. Butt, Đ. Pečarić and J. Pečarić *New Bounds For Shannon, Relative And Mandelbrot Entropies Via Hermite Interpolating Polynomial*, Demonstratio Mathematica. **51**(1) (2018), 112–130.

[17] S. I. Butt, N. Mehmood, Đ. Pečarić and J. Pečarić *New Bounds For Shannon, Relative And Mandelbrot Entropies Via Abel-Gontscharoff Interpolating Polynomial*, Math. Inequal. Appl. (submitted)

[18] S. I. Butt, N. Mehmood, Đ. Pečarić and J. Pečarić *Generalizations of Cyclic Refinements of Jensen's Inequality by Lidstone's Polynomial With Applications in Information Theory*, J. Math. Inequal. (submitted)

[19] N. Mehmood, S. I. Butt, Đ. Pečarić and J. Pečarić *New Bounds For Shannon, Relative And Mandelbrot Entropies Via Taylor's Formula*, J. Inequal. Appl. (submitted)

[20] L. Horváth, K. A. Khan and J. Pečarić, *Cyclic refinements of the discrete and integral form of Jensen's inequality with applications*, Analysis (Berlin), **36**(4) (2016), 253–263.

[21] A. Chao, L. Jost, T. C. Hsieh, K. H. Ma, W. B. Sherwin, et al. *Expected Shannon Entropy and Shannon Differentiation between Subpopulations for Neutral Genes under the Finite Island Model*, PLOS ONE, (2015) 10(6), 1–24.

[22] A. Lesne, *Shannon entropy: a rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics*, Mathematical Structures in Computer Science. (2014) 24(3), 63 pages. doi:10.1017/S0960129512000783

[23] B. R. Frieden, *Image enhancement and restoration*, Picture Processing and Digital Filtering (T. S. Huang, Editor), Springer-Verlag, Berlin (1975).

[24] S. Kullback, *Information Theory and Statistics*, J. Wiley, New York (1959).

[25] S. Kullback and R. A. Leibler, *On information and sufficiency*, Annals Math. Statist., 22, (1951), 79–86.

[26] R. M. Leahy and C. E. Goutis, *An optimal technique for constraint-based image restoration and mensuration*, IEEE Trans. on Acoustics, Speech and Signal Processing, 34, (1986), 1692–1642.

[27] 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.

[28] J. H. Justice, (editor), *Maximum Entropy and Bayssian Methods in Applied Statistics*, Cambridge University Press, Cambridge (1986).

[29] 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.

[30] C. E. Shannon, *A mathematical theory of communication*, Bull. Sept. Tech. J., 27, (1948), 370–423 and 623–656.

[31] E. Wedrowska, *Application of Kullback-Leibler Relative Entropy for Studies on the Divergence of Household Expenditures Structures*, Olsztyn Economic Journal, 6, (2011), 133–142.

[32] R. P. Agarwal and P. J. Y. Wong, *Error Inequalities in Polynomial Interpolation and their Applications*, Kluwer Academic Publishers, Dordrecht, (1993).

[33] N. Mehmood, R. P. Agarwal, S. I. Butt and J. Pečarić, *New generalizations of Popoviciu-type inequalities via new Green's functions and Montgomery identity*, J. Inequal. Appl. (2017), 2017:108.

[34] D. V. Widder, *Completely convex function and Lidstone series*, Trans. Am. Math. Soc. **51** (1942), 387–398.

[35] M. Anwar, S. Hussain, J. Pečarić, *Some inequalities for Csiszár-divergence measures*, Int. Journal of Math. Analysis 3, (2009), 1295–1304.

[36] I. Brnetić, A. K. Khuram, J. Pečarić, *Refinement of Jensen's inequality with applications to cyclic mixed symmetric means and Cauchy means*, J. Math. Inequal. 9, (2015), 1309–1321.

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

[38] I. Csiszár , J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York (1981).

[39] V. Diodato, *Dictionary of Bibliometrics*, Haworth Press, New York (1994).

[40] S. S. Dragomir, *A new refinement of Jensen's inequality in linear spaces with applications*, Math. Comput. Modelling 2, (2010), 1497–1505.

[41] S. S. Dragomir, *A refinement of Jensen's inequality with applications for f-divergence measures*, Taiwanese J. Math. 14, (2010), 153–164.

[42] L. Egghe, R. Rousseau, *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science*, Elsevier Science Publishers, New York (1990).

[43] S. Frontier, *Diversity and structure in aquatic ecosystems*, Mar. Biol. Ann. Rev. 23, (1985), 253–312.

[44] A. L. Gibbs, F. E. Su, *On choosing and bounding probability metrics*, Int. Statist. Rev. 70, (2002), 419–435.

[45] F. Liese, I. Vajda, *Convex Statistical Distances*, Teubner-Texte Zur Mathematik, vol. 95, Leipzig, Germany: Teubner (1987).

[46] B. Mandelbrot, *An informational theory of the statistical structure of language*, in Jackson, W., editor, Communication Theory, pp. 486–502, Academic Press, New York (1953).

[47] D. Mouillot, A. Lepretre, *Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity, Environmental Monitoring and Assessment*, Springer, 63, (2000), 279–295.

[48] M. E. J. Newman, *Power laws, Pareto distributions and Zipf's law*, URL: http://arxiv.org/pdf/cond-mat/0412004.pdf?origin=publication_detail (6.4.2014.)

[49] A. Rényi, *On measures of information and entropy*, Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, (1960), pp. 547–561.

[50] I. Sason, S. Verdú, *f-divergence inequalities*, IEEE Transactions on Information Theory 62, (2016), 5973–6006.

[51] I. Vajda, *Theory of Statistical Inference and Information*, Kluwer, Dordrecht (1989).

[52] J. B. Wilson, *Methods for fitting dominance/diversity curves*, J. Veget. Sci. 2, (1991), 35–46.

[53] R. W. Yeung , *Information Theory and Network Coding*, Springer (2008).

Chapter 8

Inequalities of the Jensen and Edmundson-Lah-Ribarič type for Zipf-Mandelbrot law

Rozarija Mikić, Đilda Pečarić, Josip Pečarić and Jurica Perić

Abstract. The Jensen inequality is one of the most important inequalities in modern mathematics since it implies the whole series of other classical inequalities, and one of the most famous amongst them is the so called Edmundson-Lah-Ribarič inequality. In this expository paper we start by presenting some estimates for the generalized f-divergence functional via converses of the Jensen and Edmundson-Lah-Ribarič inequalities for convex functions. We will show some Jensen and Edmundson-Lah-Ribarič type inequalities for positive linear functionals without the assumption about the convexity of the involved functions. Next, we will demonstrate some Jensen and Edmundson-Lah-Ribarič type inequalities for positive linear functionals and the class of 3-convex functions. Then, we will show how several different representations of the left side in the Edmundson-Lah-Ribarič inequality can be derived by using Hermite's interpolating polynomial written in terms of divided differences. Those representations are then utilized for obtaining different Edmundson-Lah-Ribarič type inequalities for positive linear functionals and n-convex functions. All of the mentioned general results were then used respectively to obtain the appropriate inequalities which correspond to the generalized f-divergence functional. All of the obtained results are applied to Zipf-Mandelbrot law and Zipf law in order to obtain a variety of lower and an upper bounds for different parameters. Finally, using further

generalization and improvement of Edmundson-Lah-Ribarič inequality we get some improvements of results from previous sections about generalized f-divergence functional (and also for some special cases of the function f) and Zipf-Mandelbrot law.

8.1 Introduction

Let us denote the set of all probability densities by $\mathbb{P}$, i.e. $\boldsymbol{p} = (p_1, \ldots, p_n) \in \mathbb{P}$ if $p_i \in [0,1]$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} p_i = 1$. One of the numerous applications of Probability Theory is finding an appropriate measure of distance (difference or divergence) between two probability distributions.

Consequently, many different divergence measures have been introduced and extensively studied, for example Kullback-Leibler divergence, Hellinger divergence, Renyi divergence, Bhattacharyya divergence, harmonic divergence, Jeffreys divergence, triangular divergence etc. All of the mentioned divergences are special cases of Csiszár f-divergence.

These measures of distance between two probability distributions have an important application in a great number of fields such as: anthropology, genetics, economics and political science, biology, approximation of probability distributions ([7], [22]), signal processing ([18]) and pattern recognition ([2], [4]), analysis of contingency tables ([8]), ecological studies, music etc.

A large number of papers has been written on the subject of inequalities for different types of divergences. Since the functions that are used to define most of the divergences are convex, Jensen's inequality and its converses play an important role in the mentioned inequalities.

Csiszár [8]-[9] introduced the $f-$divergence functional as

$$D_f(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right), \tag{8.1}$$

where $f\colon [0,+\infty\rangle$ is a convex function, and it represent a "distance function" on the set of probability distributions $\mathbb{P}$.

Dragomir [11] gave the following upper bound for the Csiszár divergence functional

$$D_f(\boldsymbol{p},\boldsymbol{q}) \le \frac{M-1}{M-m} f(m) + \frac{1-m}{M-m} f(M), \tag{8.2}$$

where f is a convex function on the interval $[m,M]$, $\boldsymbol{p} = (p_1, \ldots, p_n), \boldsymbol{q} = (q_1, \ldots, q_n) \in \mathbb{P}$ and $m \le p_i/q_i \le M$ for every $i = 1, \ldots, n$ (then it easily follows that $1 \in [m,M]$).

The Kullback-Leibler divergence, also called relative entropy or KL divergence

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

is a measure of the non-symmetric difference between two probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$, but it is not a true metric because it does not obey the triangle inequality and in general

$D_{KL}(\boldsymbol{p},\boldsymbol{q}) \neq D_{KL}(\boldsymbol{q},\boldsymbol{p})$. The Kullback-Leibler divergence was introduced by Kullback and Leibler in [21], and it is a special case of the Csiszár divergence for $f(t) = t\log t$.

Let $\boldsymbol{p} = (p_1,\dots,p_n)$ and $\boldsymbol{p} = (q_1,\dots,q_n)$ be probability distributions and let $[m,M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$ and $p_i/q_i \in [m,M]$ for every $i = 1,\dots,n$. Apart from the Kullback-Leibler divergence, some other well-known divergences that are special cases of f-divergence for different choices of the function f are as follows.

▷ **Hellinger divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

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

with the corresponding generating function $f(t) = \frac{1}{2}(1-\sqrt{t})^2, t > 0$.

▷ **Renyi divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_\alpha(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n} q_i^{\alpha-1}p_i^{\alpha},\ \alpha \in \mathbb{R},$$

▷ **Harmonic divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_{Ha}(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n} \frac{2p_iq_i}{p_i+q_i},$$

and the corresponding generating function $f(t) = \frac{2t}{1+t}$.

▷ **Jeffreys divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_J(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{2}\sum_{i=1}^{n}(q_i - p_i)\log\frac{q_i}{p_i},$$

with the corresponding generating function $f(t) = (1-t)\log\frac{1}{t}, t > 0$.

In order to use nonnegative probability distributions in the f-divergence functional, Horvath et. al. in [15] defined

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

and gave the following definition of a generalized f-divergence functional.

Definition 8.1 *Let $J \subset \mathbb{R}$ be an interval, and let $f\colon J \to \mathbb{R}$ be a function. Let $P = (p_1,\dots,p_n)$ be an n-tuple of real numbers and $Q = (q_1,\dots,q_n)$ be an n-tuple of nonnegative real numbers such that $p_i/q_i \in J$ for every $i = 1,\dots,n$. Then let*

$$\hat{D}_f(\boldsymbol{p},\boldsymbol{q}) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right). \tag{8.3}$$

The aim of this chapter is to give some inequalities of the Jensen and Edmundson-Lah-Ribarič type for generalized f-divergence functional, and then exploit them in obtaining various bounds for the difference between famous Zipf-Mandelbrot laws with different parameters. It is organised in the following manner: in Section 8.2 we will state some general results for positive linear functionals concerning different inequalities of the Jensen and Edmundson-Lah-Ribarič type that we will use in the proofs of our results; in Section 8.3 we give our results concerning the generalized f-divergence functional for different classes of functions, not only the convex ones, and finally in Section 8.4 we derive diverse bounds for the divergence of Zipf and Zipf-Mandelbrot law. Finally in Section 8.5 we give some furher improvements of results for generalized f-diveregence and for diveregence of Zipf-Mandelbrot law.

8.2 Preliminaries

Let E be a non-empty set and L a vector space of real functions $f\colon E \to \mathbb{R}$ with the following properties:

(L1): $f,g \in L \Rightarrow (af+bg) \in L$ for all $a,b \in \mathbb{R}$;

(L2): $\boldsymbol{1} \in L$, that is, if $f(t) = 1$ for every $t \in E$, then $f \in L$.

(L3): if $f,g \in L$, then $\min\{f,g\} \in L$ or $\max\{f,g\} \in L$.

We say that $A\colon L \to \mathbb{R}$ is a positive linear functional if:

(A1): $A(af+bg) = aA(f) + bA(g)$ for $f,g \in L$ and $a,b \in \mathbb{R}$;

(A2): $f \in L$, $f(t) \geq 0$ for every $t \in E \Rightarrow A(f) \geq 0$.

We say that a functional A is normalized if $A(\boldsymbol{1}) = 1$.

Throughout this chapter, if a function is defined on an interval $[m,M]$ without any further emphasis we assume that the bounds of that interval are finite.

Jessen [17] gave the following generalization of Jensen's inequality for convex functions (see also [15, str.47]):

Theorem 8.1 *([17]) Let L be a vector space of real functions defined on an non-empty set E that has properties $(L1)$ and $(L2)$, and let us assume that ϕ is a continuous convex function on an interval $I \subset \mathbb{R}$. If A is a normalized positive linear functional, then for every $f \in L$ such that $\phi(f) \in L$ we have $A(f) \in I$ and*

$$\phi(A(f)) \leq A(\phi(f)). \tag{8.4}$$

Next result is a generalization of the Edmundson-Lah-Ribarič inequality for linear functionals and it was proved by Beesack and Pečarić in [3] (see also [15, str.98]):

Theorem 8.2 *([3]) Let ϕ be a convex function on $I = [m,M]$, let L be a vector space of real functions defined on an non-empty set E that has properties $(L1)$ and $(L2)$, and let A be a normalized positive linear functional. Then for every $f \in L$ such that $\phi(f) \in L$ (so $m \le f(t) \le M$ for all $t \in E$), we have*

$$A(\phi(f)) \le \frac{M - A(f)}{M - m}\phi(m) + \frac{A(f) - m}{M - m}\phi(M). \tag{8.5}$$

Klaričić Bakula, Pečarić and Perić in [19] gave the following improvement of the inequality (11.3.1).

Theorem 8.3 *([19]) Let L be a vector space of real functions defined on an non-empty set E that has properties $(L1)$, $(L2)$ and $(L3)$ and let A be a normalized positive linear functional on L. If ϕ is a convex function on $[m,M]$, then for every $f \in L$ such that $\phi(f) \in L$ we have $A(f) \in [m,M]$ and*

$$A(\phi(f)) \le \frac{M - A(f)}{M - m}\phi(m) + \frac{A(f) - m}{M - m}\phi(M) - A(\tilde{f})\delta_\phi, \tag{8.6}$$

where $\tilde{f}$ and δ_ϕ are defined in

$$\tilde{f} = \frac{1}{2}\boldsymbol{1} - \frac{1}{M - m}\left| f - \frac{m + M}{2}\boldsymbol{1} \right|, \quad \delta_\phi = \phi(m) + \phi(M) - 2\phi\left(\frac{m + M}{2}\right). \tag{8.7}$$

The subsequent two result provide us with bounds for the difference in the Jensen and Edmundson-Lah-Ribarič inequalities respectively.

Theorem 8.4 ([20]) *Let ϕ be a continuous convex function on the interval I whose interior contains interval $[m,M]$, let L be a vector space of real functions defined on a non-empty set E such that it has properties $(L1)$, $(L2)$ and $(L3)$. Let A be any normalized positive linear functional on L. Then for every function $f \in L$ such that $\phi(f) \in L$ and which satisfies the bounds $m \le f(t) \le M$ for every $t \in E$ we have*

$$\begin{aligned} 0 &\le A(\phi(f)) - \phi(A(f)) \\ &\le (M - A(f))(A(f) - m) \sup_{t \in \langle m,M \rangle} \Psi_\phi(t;m,M) - A(\tilde{f})\delta_\phi \\ &\le (M - A(f))(A(f) - m)\frac{\phi'_-(M) - \phi'_+(m)}{M - m} - A(\tilde{f})\delta_\phi \\ &\le \frac{1}{4}(M - m)(\phi'_-(M) - \phi'_+(m)) - A(\tilde{f})\delta_\phi. \end{aligned} \tag{8.8}$$

We also have

$$\begin{aligned} 0 &\le A(\phi(f)) - \phi(A(f)) \le \frac{1}{4}(M - m)^2\Psi_\phi(A(f);m,M) - A(\tilde{f})\delta_\phi \\ &\le \frac{1}{4}(M - m)(\phi'_-(M) - \phi'_+(m)) - A(\tilde{f})\delta_\phi, \end{aligned} \tag{8.9}$$

where $\tilde{f}$ and δ_ϕ are defined in (8.7), $\Psi_\phi(\cdot;m,M)\colon \langle m,M\rangle \to \mathbb{R}$ is defined by

$$\Psi_\phi(t;m,M) = \frac{1}{M-m}\Big(\frac{\phi(M)-\phi(t)}{M-t} - \frac{\phi(t)-\phi(m)}{t-m}\Big), \tag{8.10}$$

and we assume that $\Psi_\phi(f;m,M) \in L$. If ϕ is concave on I, then the inequality signs are reversed.

Theorem 8.5 ([20]) *Let ϕ be a continuous convex function on the interval I whose interior contains interval $[m,M]$, let L be a vector space of real functions defined on a nonempty set E such that it has properties $(L1)$, $(L2)$ and $(L3)$. Let A be any normalized positive linear functional on L. Then for every function $f \in L$ such that $\phi(f) \in L$ and which satisfies the bounds $m \le f(t) \le M$ for every $t \in E$ we have the following sequences of inequalities*

(i)

$$\begin{aligned}
0 &\le \frac{A(f)-m}{M-m}\phi(M) + \frac{M-A(f)}{M-m}\phi(m) - A(\phi(f)) - A(\tilde{f})\delta_\phi \\
&\le A[(M-f)(f-m)] \sup_{t\in\langle m,M\rangle} \Psi_\phi(t;m,M) - A(\tilde{f})\delta_\phi \\
&\le \frac{A[(M-f)(f-m)]}{M-m}(\phi'_-(M) - \phi'_+(m)) - A(\tilde{f})\delta_\phi \\
&\le \frac{(M-A(f))(A(f)-m)}{M-m}(\phi'_-(M) - \phi'_+(m)) - A(\tilde{f})\delta_\phi \\
&\le \frac{1}{4}(M-m)(\phi'_-(M) - \phi'_+(m)) - A(\tilde{f})\delta_\phi
\end{aligned} \tag{8.11}$$

(ii)

$$\begin{aligned}
0 &\le \frac{A(f)-m}{M-m}\phi(M) + \frac{M-A(f)}{M-m}\phi(m) - A(\phi(f)) - A(\tilde{f})\delta_\phi \\
&\le A[(M-f)(f-m)] \sup_{t\in\langle m,M\rangle} \Psi_\phi(t;m,M) - A(\tilde{f})\delta_\phi \\
&\le (M-A(f))(A(f)-m) \sup_{t\in\langle m,M\rangle} \Psi_\phi(t;m,M) - A(\tilde{f})\delta_\phi \\
&\le \frac{(M-A(f))(A(f)-m)}{M-m}(\phi'_-(M) - \phi'_+(m)) - A(\tilde{f})\delta_\phi \\
&\le \frac{1}{4}(M-m)(\phi'_-(M) - \phi'_+(m)) - A(\tilde{f})\delta_\phi
\end{aligned} \tag{8.12}$$

(iii)

$$\begin{aligned}
0 &\le \frac{A(f)-m}{M-m}\phi(M) + \frac{M-A(f)}{M-m}\phi(m) - A(\phi(f)) - A(\tilde{f})\delta_\phi \\
&\le \frac{1}{4}(M-m)^2 A(\Psi_\phi(f;m,M)) - A(\tilde{f})\delta_\phi \\
&\le \frac{1}{4}(M-m)(\phi'_-(M) - \phi'_+(m)) - A(\tilde{f})\delta_\phi
\end{aligned} \tag{8.13}$$

where $\tilde{f}$ and δ_ϕ are defined in (8.7), and $\Psi_\phi(\cdot;m,M)$ is defined in (8.10). If the function ϕ is concave, then the inequality signs are reversed.

The following results give us a class of inequalities of the Jensen and Edmundson-Lah-Ribarič type which are valid for functions with bounded second order divided differences, and are obtained in paper [26]. This was a significant improvement compared to the results from above, because these hold for a much wider class of functions than the class of convex functions.

Theorem 8.6 ([26]) Let ϕ be a function on an interval of real numbers $[m,M]$ such that there exist $\gamma,\Gamma \in \mathbb{R}$ such that $\gamma \leq [m,t,M]\phi \leq \Gamma$ holds for every $t \in [m,M]$, that is, such that its second order divided difference in m,t and M is bounded for every $t \in [m,M]$. Let L satisfy conditions (L1) and (L2) on E and let A be any positive linear functional on L with $A(\boldsymbol{1}) = 1$. Then

$$\begin{aligned}\gamma A\left[(M\boldsymbol{1}-f)(f-m\boldsymbol{1})\right] &\leq \frac{M-A(f)}{M-m}\phi(m)+\frac{A(f)-m}{M-m}\phi(M)-A\left(\phi(f)\right)\\ &\leq \Gamma A\left[(M\boldsymbol{1}-f)(f-m\boldsymbol{1})\right] \end{aligned} \tag{8.14}$$

holds for any $f \in L$ such that $\phi \circ f \in L$.

Remark 8.1 ([26]) There are two more cases that need to be considered.

- If $0 \leq \gamma < \Gamma < \infty$, then the function ϕ is convex, so we have that

$$\begin{aligned}0 \leq &\gamma A\left[(M\boldsymbol{1}-f)(f-m\boldsymbol{1})\right]\\ &\frac{M-A(f)}{M-m}\phi(m)+\frac{A(f)-m}{M-m}\phi(M)-A\left(\phi(f)\right)\\ \leq &\Gamma A\left[(M\boldsymbol{1}-f)(f-m\boldsymbol{1})\right]\\ \leq &\Gamma(M-A(f))(A(f)-m) \leq \frac{\Gamma}{4}(M-m)^2\end{aligned} \tag{8.15}$$

  holds for any $f \in L$ such that $\phi \circ f \in L$.

- If $-\infty < \gamma < \Gamma \leq 0$, then the function ϕ is concave, so we have that

$$\begin{aligned}\leq &\frac{\gamma}{4}(M-m)^2 \leq \gamma(M-A(f))(A(f)-m)\\ \leq &\gamma A\left[(M\boldsymbol{1}-f)(f-m\boldsymbol{1})\right]\\ \leq &\frac{M-A(f)}{M-m}\phi(m)+\frac{A(f)-m}{M-m}\phi(M)-A\left(\phi(f)\right)\\ \leq &\Gamma A\left[(M\boldsymbol{1}-f)(f-m\boldsymbol{1})\right] \leq 0\end{aligned} \tag{8.16}$$

  holds for any $f \in L$ such that $\phi \circ f \in L$.

Theorem 8.7 ([26]) Let ϕ be a function on an interval of real numbers $[m,M]$ such that there exist $\gamma,\Gamma \in \mathbb{R}$ such that $\gamma \leq [m,t,M]\phi \leq \Gamma$ holds for every $t \in [m,M]$, that is, such that its second order divided difference in m,t and M is bounded for every $t \in [m,M]$. Let L satisfy conditions (L1) and (L2) on E and let A be any positive linear functional on L with $A(\boldsymbol{1}) = 1$. Then

$$\begin{aligned} &\gamma(M - A(f))(A(f) - m) - \Gamma A\left[(M\boldsymbol{1} - f)(f - m\boldsymbol{1})\right] \\ \leq &A\left(\phi(f)\right) - \phi(A(f)) \leq \Gamma(M - A(f))(A(f) - m) - \gamma A\left[(M\boldsymbol{1} - f)(f - m\boldsymbol{1})\right] \end{aligned} \tag{8.17}$$

holds for any $f \in L$ such that $\phi \circ f \in L$.

The results that follow represent different classes of inequalities of the Jensen and Edmundson-Lah-Ribarič type that hold for n-convex functions, so first we need to recall some definitions and properties.

Definition of the n-convex function is characterized by nth-order divided difference. The nth-order divided difference of a function $f\colon [a,b] \to \mathbb{R}$ at mutually distinct points $t_0, t_1, \ldots, t_n \in [a,b]$ is defined recursively by

$$\begin{aligned} [t_i]f &= f(t_i), \quad i = 0, \ldots, n, \\ [t_0, \ldots, t_n]f &= \frac{[t_1, \ldots, t_n]f - [t_0, \ldots, t_{n-1}]f}{t_n - t_0}. \end{aligned}$$

The value $[t_0, \ldots, t_n]f$ is independent of the order of the points $t_0, \ldots, t_n$. Definition of divided differences can be extended to include the cases in which some or all the points coincide (see e.g. [1], [15]):

$$f[\underbrace{a, \ldots, a}_{n\ times}] = \frac{1}{(n-1)!} f^{(n-1)}(a), \quad n \in \mathbb{N}.$$

Regarding third order divided differences, in the case in which some or all the points coincide they are defined in the following way.

- If the function f is differentiable on $[a,b]$ and $t, t_0, t_1 \in [a,b]$ are mutually different points, then

$$\begin{aligned} [t,t,t_0,t_1]f = &\frac{f'(t)}{(t-t_0)(t-t_1)} + \frac{f(t)(t_0+t_1-2t)}{(t-t_0)^2(t-t_1)^2} \\ &+ \frac{f(t_0)}{(t_0-t)^2(t_0-t_1)} + \frac{f(t_1)}{(t_1-t)^2(t_1-t_0)}. \end{aligned} \tag{8.18}$$

- If the function f is differentiable on $[a,b]$ and $t, t_0 \in [a,b]$ are mutually different points, then

$$[t,t,t_0,t_0]f = \frac{1}{(t_0-t)^3}\left[(t_0-t)(f'(t_0)+f'(t)) + 2(f(t)-f(t_0))\right]. \tag{8.19}$$

- If the function f is twice differentiable on $[a,b]$ and $t, t_0 \in [a,b]$ are mutually different points, then

$$[t,t,t,t_0]f = \frac{1}{(t_0-t)^3}\left[f(t_0) - \sum_{k=0}^{2} \frac{f^{(k)}(t)}{k!}(t_0-t)^k\right]. \tag{8.20}$$

- If the function f is three times differentiable on $[a,b]$ and $t \in [a,b]$, then

$$[t,t,t,t]f = \frac{f'''(t)}{3!}. \tag{8.21}$$

A function $f\colon [m,M] \to \mathbb{R}$ is said to be n-convex ($n \geq 0$) if and only if for all choices of $(n+1)$ distinct points $t_0, t_1, \ldots, t_n \in [m,M]$, we have $[t_0, \ldots, t_n]f \geq 0$.

We can extend the definition of 3-convex functions by including the cases in which some or all of the points coincide. This is given in the following theorem which can be easily proven by using the mean value theorem for divided differences (see e.g. [16]).

Theorem 8.8 *Let a function f be defined on an interval $I \subseteq \mathbb{R}$. The following equivalences hold.*

(i) If $f \in \mathscr{C}(I)$, then f is 3-convex if and only if $[t,t,t_0,t_1]f \geq 0$ for all mutually different points $t, t_0, t_1 \in I$.

(ii) If $f \in \mathscr{C}(I)$, then f is 3-convex if and only if $[t,t,t_0,t_0]f \geq 0$ for all mutually different points $t, t_0 \in I$.

(iii) If $f \in \mathscr{C}^2(I)$, then f is 3-convex if and only if $[t,t,t,t_0]f \geq 0$ for all mutually different points $t, t_0 \in I$.

(iv) If $f \in \mathscr{C}^3(I)$, then f is 3-convex if and only if $[t,t,t,t]f \geq 0$ for every $t \in I$.

The first result of this type is a reversed Edmundson-Lah-Ribarič inequality for 3-convex functions.

Theorem 8.9 ([24]) Let L satisfy conditions (L1) and (L2) on a non-empty set E and let A be any positive linear functional on L with $A(\boldsymbol{1}) = 1$. Let ϕ be a 3-convex function on an interval of real numbers I whose interior contains the interval $[m,M]$. Then

$$\begin{aligned} &\frac{A\left[(M\boldsymbol{1}-f)(f-m\boldsymbol{1})\right]}{M-m}\left(\frac{\phi(M)-\phi(m)}{M-m}-\phi'_+(m)\right) \\ \leq &\frac{M-A(f)}{M-m}\phi(m)+\frac{A(f)-m}{M-m}\phi(M)-A(\phi(f)) \\ \leq &\frac{A\left[(M\boldsymbol{1}-f)(f-m\boldsymbol{1})\right]}{M-m}\left(\phi'_-(M)-\frac{\phi(M)-\phi(m)}{M-m}\right) \end{aligned} \tag{8.22}$$

holds for any $f \in L$ such that $\phi \circ f \in L$ and $m \leq f(t) \leq M$ for $t \in E$. If the function $-\phi$ is 3-convex, then the inequalities are reversed.

Next result is a Jensen-type inequality for 3-convex functions.

Theorem 8.10 ([24]) Let L satisfy conditions (L1) and (L2) on a non-empty set E and let A be any positive linear functional on L with $A(\boldsymbol{1}) = 1$. Let ϕ be a 3-convex function on an interval of real numbers I whose interior contains the interval $[m,M]$. Then

$$\frac{(M-A(f))(A(f)-m)}{M-m}\left(\frac{\phi(M)-\phi(m)}{M-m}-\phi'_+(m)\right)$$

$$
\begin{aligned}
&-\frac{A\left[(M\mathbf{1}-f)(f-m\mathbf{1})\right]}{M-m}\left(\phi_-'(M)-\frac{\phi(M)-\phi(m)}{M-m}\right) \\
\leq &A\left(\phi(f)\right)-\phi(A(f)) \leq \frac{(M-A(f))(A(f)-m)}{M-m}\left(\phi_-'(M)-\frac{\phi(M)-\phi(m)}{M-m}\right) \\
&-\frac{A\left[(M\mathbf{1}-f)(f-m\mathbf{1})\right]}{M-m}\left(\frac{\phi(M)-\phi(m)}{M-m}-\phi_+'(m)\right)
\end{aligned} \tag{8.23}
$$

holds for any $f \in L$ such that $\phi \circ f \in L$ and $m \leq f(t) \leq M$ for $t \in E$. If the function $-\phi$ is 3-convex, then the inequalities are reversed.

A different Edmundson-Lah-Ribarič type inequalities for 3-convex functions are given in the following theorems.

Theorem 8.11 ([27]) *Let L satisfy conditions (L1) and (L2) on a non-empty set E and let A be any positive linear functional on L with $A(\mathbf{1}) = 1$. Let ϕ be a 3-convex function defined on an interval of real numbers I whose interior contains the interval $[m,M]$ and differentiable on $\langle m,M\rangle$. Then*

$$
\begin{aligned}
&(A(f)-m)\left[\frac{\phi(M)-\phi(m)}{M-m}-\frac{\phi_+'(m)}{2}\right]-\frac{1}{2}A[(f-m\mathbf{1})\phi'(f)] \\
\leq &\frac{M-A(f)}{M-m}\phi(m)+\frac{A(f)-m}{M-m}\phi(M)-A(\phi(f)) \\
\leq &\frac{1}{2}A[(M\mathbf{1}-f)\phi'(f)]-(M-A(f))\left[\frac{\phi(M)-\phi(m)}{M-m}-\frac{\phi_-'(M)}{2}\right]
\end{aligned} \tag{8.24}
$$

holds for any $f \in L$ such that $\phi \circ f \in L$ and $m \leq f(t) \leq M$ for $t \in E$. If the function $-\phi$ is 3-convex, then the inequalities are reversed.

Theorem 8.12 ([27]) *Let L satisfy conditions (L1) and (L2) on a non-empty set E and let A be any positive linear functional on L with $A(\mathbf{1}) = 1$. Let ϕ be a 3-convex function defined on an interval of real numbers I whose interior contains the interval $[m,M]$ and differentiable on $\langle m,M\rangle$. Then*

$$
\begin{aligned}
&(M-A(f))\left[\phi_-'(M)-\frac{\phi(M)-\phi(m)}{M-m}\right]-\frac{\phi_-''(M)}{2}A[(M\mathbf{1}-f)^2] \\
\leq &\frac{M-A(f)}{M-m}\phi(m)+\frac{A(f)-m}{M-m}\phi(M)-A(\phi(f)) \\
\leq &(A(f)-m)\left[\frac{\phi(M)-\phi(m)}{M-m}-\phi_+'(m)\right]-\frac{\phi_+''(m)}{2}A[(f-m\mathbf{1})^2]
\end{aligned} \tag{8.25}
$$

holds for any $f \in L$ such that $\phi \circ f \in L$ and $m \leq f(t) \leq M$ for $t \in E$. If the function $-\phi$ is 3-convex, then the inequalities are reversed.

Since all of the results below are obtained by using Hermite's interpolating polynomials in terms of divided differences, let us introduce the following notations. For a given function $f\colon [m,M] \to \mathbb{R}$ denote:

$$LR(f,g,m,M,A) = A(f(g)) - \frac{M-A(g)}{M-m} f(m) - \frac{A(g)-m}{M-m} f(M) \tag{8.26}$$

and

$$R_v(t) = (t-m)^v (t-M)^{n-v} f[t;\underbrace{m,\ldots,m}_{v\ times};\underbrace{M,M,\ldots,M}_{(n-v)\ times}]. \tag{8.27}$$

Theorem 8.13 ([28]) Let L satisfy conditions (L1) and (L2) on a non-empty set E and let A be any positive linear functional on L with $A(\boldsymbol{1}) = 1$. Let $f \in \mathscr{C}^n([m,M])$, and let $g \in L$ be any function such that $f \circ g \in L$. If the function f is n-convex and if n and $v \geq 3$ are of different parity, then

$$\begin{aligned} LR(f,g,m,M,A) &\leq (A(g)-m)\,(f[m,m]-f[m,M]) + \sum_{k=2}^{v-1} \frac{f^{(k)}(m)}{k!} A\left[(g-m\boldsymbol{1})^k\right] \\ &+ \sum_{k=1}^{n-v} f[\underbrace{m,\ldots,m}_{v\ times};\underbrace{M,\ldots,M}_{k\ times}] A\left[(g-m\boldsymbol{1})^v (g-M\boldsymbol{1})^{k-1}\right]. \end{aligned} \tag{8.28}$$

Inequality (8.28) also holds when the function f is n-concave and n and v are of equal parity. In case when the function f is n-convex and n and v are of equal parity, or when the function f is n-concave and n and v are of different parity, the inequality sign in (8.28) is reversed.

Theorem 8.14 ([28]) Let L satisfy conditions (L1) and (L2) on a non-empty set E and let A be any positive linear functional on L with $A(\boldsymbol{1}) = 1$. Let $f \in \mathscr{C}^n([m,M])$, and let $g \in L$ be any function such that $f \circ g \in L$. If the function f is n-convex and if $v \geq 3$ is odd, then

$$\begin{aligned} LR(f,g,m,M,A) &\leq (M-A(g))\,(f[m,M]-f[M,M]) + \sum_{k=2}^{v-1} \frac{f^{(k)}(M)}{k!} A[(g-M\boldsymbol{1})^k] \\ &+ \sum_{k=1}^{n-v} f[\underbrace{M,\ldots,M}_{v\ times};\underbrace{m,\ldots,m}_{k\ times}] A[(g-M\boldsymbol{1})^v (g-m\boldsymbol{1})^{k-1}] \end{aligned} \tag{8.29}$$

Inequality (8.29) also holds when the function f is n-concave and v is even. In case when the function f is n-convex and v is even, or when the function f is n-concave and v is odd, the inequality sign in (8.29) is reversed.

Theorem 8.15 ([28]) Let L satisfy conditions (L1) and (L2) on a non-empty set E and let A be any positive linear functional on L with $A(\boldsymbol{1}) = 1$. Let $f \in \mathscr{C}^n([m,M])$, and let $g \in L$ be any function such that $f \circ g \in L$. If the function f is n-convex and if n is odd, then

$$\sum_{k=2}^{n-1} f[m;\underbrace{M,\ldots,M}_{k\ times}] A\left[(g-m\boldsymbol{1})(g-M\boldsymbol{1})^{k-1}\right] \leq LR(f,g,m,M,A) \tag{8.30}$$

$$\leq f[m,m;b]A[(g-m\mathbf{1})(g-M\mathbf{1})]+\sum_{k=2}^{n-2} f[m,m;\underbrace{M,\dots,M}_{k\ times}]A\left[(g-m\mathbf{1})^2(g-M\mathbf{1})^{k-1}\right].$$

Inequalities (8.30) also hold when the function f is n-concave and n is even. In case when the function f is n-convex and n is even, or when the function f is n-concave and n is odd, the inequality signs in (8.30) are reversed.

Theorem 8.16 ([28]) Let L satisfy conditions (L1) and (L2) on a non-empty set E and let A be any positive linear functional on L with $A(\mathbf{1}) = 1$. Let $f \in \mathscr{C}^n([m,M])$, and let $g \in L$ be any function such that $f \circ g \in L$. If the function f is n-convex, then

$$f[M,M;m]A[(g-M\mathbf{1})(g-m\mathbf{1})]+\sum_{k=2}^{n-2} f[M,M;\underbrace{m,\dots,m}_{k\ times}]A[(g-M\mathbf{1})^2(g-m\mathbf{1})^{k-1}]$$

$$\leq LR(f,g,m,M,A) \leq \sum_{k=1}^{n-1} f[M;\underbrace{m,\dots,m}_{k\ times}]A[(g-M\mathbf{1})(g-m\mathbf{1})^{k-1}]. \tag{8.31}$$

If the function f is n-concave, the inequality signs in (8.31) are reversed.

8.3 Inequalities for generalized f-divergence

Our first result in this section is an improved version of Dragomir's result (8.2) for the generalized f-divergence functional, and it provides us an upper bound for the mentioned functional.

Theorem 8.17 ([25]) *Let $[m,M] \subset \mathbb{R}$ be an interval, let $f\colon [m,M] \to \mathbb{R}$ be a function and let δ_f be defined in (8.7). Let $\boldsymbol{p} = (p_1,\dots,p_n)$ be an n-tuple of real numbers and $\boldsymbol{q} = (q_1,\dots,q_n)$ be an n-tuple of nonnegative real numbers such that $p_i/q_i \in [m,M]$ for every $i = 1,\dots,n$. If the function f is convex, we have*

$$\begin{aligned}\hat{D}_f(\boldsymbol{p},\boldsymbol{q}) \leq{}& \frac{MQ_n-P_n}{M-m}f(m)+\frac{P_n-mQ_n}{M-m}f(M)\\ &-\left(\frac{Q_n}{2}-\frac{1}{M-m}\sum_{i=1}^{n}\left|p_i-\frac{m+M}{2}q_i\right|\right)\delta_f,\end{aligned} \tag{8.32}$$

where $P_n = \sum_{i=1}^n p_i$ and $Q_n = \sum_{i=1}^n q_i$. If the function f is concave, then the inequality sign is reversed.

Proof. Let $f\colon [m,M] \to \mathbb{R}$ be a convex function. For an n-tuple of real numbers $\boldsymbol{x} = (x_1,\dots,x_n)$, an n-tuple of positive numbers $\boldsymbol{p} = (p_1,\dots,p_n)$ and a normalized positive linear functional $A(\boldsymbol{x}) = \frac{1}{P_n}\sum_{i=1}^n p_i x_i$, from Theorem 8.3 we have

$$\frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) \leq \frac{M-\overline{x}}{M-m}f(m)+\frac{\overline{x}-m}{M-m}f(M)$$

$$-\frac{1}{P_n}\sum_{i=1}^{n} p_i\left(\frac{1}{2}-\frac{1}{M-m}\left|x_i-\frac{m+M}{2}\right|\right)\delta_f, \tag{8.33}$$

where $\bar{x}=\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i$. Since $\boldsymbol{q}=(q_1,\ldots,q_n)$ are nonnegative real numbers, we can put

$$p_i = q_i \text{ and } x_i = \frac{p_i}{q_i}$$

in (8.33) and get

$$\begin{aligned}\frac{1}{Q_n}\sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) &\le \frac{M-\frac{1}{Q_n}\sum_{i=1}^{n} q_i\frac{p_i}{q_i}}{M-m}f(m)+\frac{\frac{1}{Q_n}\sum_{i=1}^{n} q_i\frac{p_i}{q_i}-m}{M-m}f(M)\\ &\quad-\frac{1}{Q_n}\left(\frac{Q_n}{2}-\frac{1}{M-m}\sum_{i=1}^{n} q_i\left|\frac{p_i}{q_i}-\frac{m+M}{2}\right|\right)\delta_f,\end{aligned}$$

and after multiplying by Q_n we get (8.32). □

Remark 8.2 From $m \le p_i/q_i \le M$ it easily follows that (see [19])

$$-\frac{M-m}{2}q_i \le p_i-\frac{m+M}{2}q_i \le \frac{M-m}{2}q_i, \text{ i.e. } \left|p_i-\frac{m+M}{2}q_i\right| \le \frac{M-m}{2}q_i$$

which together with $\delta_f \ge 0$ for a convex function f gives us

$$\left(\frac{Q_n}{2}-\frac{1}{M-m}\sum_{i=1}^{n}\left|p_i-\frac{m+M}{2}q_i\right|\right)\delta_f \ge 0.$$

Remark 8.3 If in the previous theorem we take $\boldsymbol{p}$ and $\boldsymbol{q}$ to be probability distributions, we directly get an improvement of Dragomir's result for the Csiszár f-divergence functional:

$$\begin{aligned}D_f(\boldsymbol{p},\boldsymbol{q}) &\le \frac{M-1}{M-m}f(m)+\frac{1-m}{M-m}f(M)\\ &\quad-\left(\frac{1}{2}-\frac{1}{M-m}\sum_{i=1}^{n}\left|p_i-\frac{m+M}{2}q_i\right|\right)\delta_f.\end{aligned}$$

Next result is a special case of Theorem 8.17, and provides us with bounds for the Kullback-leibler divergence of two probability distributions.

Corollary 8.1 ([25]) *Let $[m,M]\subset\mathbb{R}$ be an interval and let us assume that the base of the logarithm is greater than* 1.

- *Let $\boldsymbol{p}=(p_1,\ldots,p_n)$ and $\boldsymbol{q}=(q_1,\ldots,q_n)$ be n-tuples of nonnegative real numbers such that $p_i/q_i\in[m,M]$ for every $i=1,\ldots,n$. Then*

$$\begin{aligned}\sum_{i=1}^{n} p_i\log\left(\frac{p_i}{q_i}\right) &\le Q_n\frac{Mm}{M-m}\log\left(\frac{m}{M}\right)+\frac{P_n}{M-m}\log\left(\frac{M^M}{m^m}\right) \qquad (8.34)\\ &\quad-\left(\frac{Q_n}{2}-\frac{1}{M-m}\sum_{i=1}^{n}\left|p_i-\frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M}+M\log\frac{2M}{m+M}\right).\end{aligned}$$

- *Let* $\boldsymbol{p} = (p_1, \ldots, p_n)$ *and* $\boldsymbol{q} = (q_1, \ldots, q_n) \in \mathbb{P}$ *be probability distributions such that* $m \le p_i/q_i \le M$ *holds for every* $i = 1, \ldots, n$. *Then*

$$\begin{aligned} D_{KL}(\boldsymbol{p},\boldsymbol{q}) &\le \frac{Mm}{M-m}\log\left(\frac{m}{M}\right) + \frac{1}{M-m}\log\left(\frac{M^M}{m^m}\right) \\ &- \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right). \end{aligned} \tag{8.35}$$

If the base of the logarithm is less than 1, *the inequality sign in the inequalities above is reversed.*

Proof. Let $\boldsymbol{p} = (p_1, \ldots, p_n)$ and $\boldsymbol{q} = (q_1, \ldots, q_n)$ be an n-tuples of nonnegative real numbers. Since the function $t \mapsto t\log t$ is convex when the base of the logarithm is greater than 1, the inequality (8.34) follows from Theorem 8.17, inequality (8.17), by setting $f(t) = t\log t$.

Inequality (8.35) is a special case of the inequality (8.34) for probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$. □

Next result is obtained by utilizing Theorem 8.4, and it also gives us bounds for the generalized f-divergence functional. Concurrently, it represents an improvement of bounds for f-divergence functional obtained by Dragomir in the paper [11].

Theorem 8.18 ([25]) *Let* $I \subset \mathbb{R}$ *be an interval such that its interior contains the interval* $[m,M]$, *let* $f\colon I \to \mathbb{R}$ *be a continuous function and let* δ_f *be defined in (8.7). Let* $\boldsymbol{p} = (p_1, \ldots, p_n)$ *be an n-tuple of real numbers and* $\boldsymbol{q} = (q_1, \ldots, q_n)$ *be an n-tuple of nonnegative real numbers such that* $p_i/q_i \in [m,M]$ *for every* $i = 1, \ldots, n$. *Let* Ψ_f *be defined in (8.10). If the function* f *is convex, then*

$$\begin{aligned} 0 &\le \hat{D}_f(\boldsymbol{p},\boldsymbol{q}) - Q_n f\left(\frac{P_n}{Q_n}\right) \\ &\le Q_n\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right)\sup_{t\in\langle m,M\rangle}\Psi_f(t;m,M) \\ &\quad - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\delta_f \\ &\le \frac{Q_n}{M-m}\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right)\left(f'_-(M) - f'_+(m)\right) \\ &\quad - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\delta_f \\ &\le \frac{Q_n}{4}(M-m)(f'_-(M) - f'_+(m)) - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\delta_f. \end{aligned} \tag{8.36}$$

If the function f *is concave, the inequality signs are reversed.*

Proof. Let $f\colon [m,M] \to \mathbb{R}$ be a convex function. Let $\boldsymbol{x} = (x_1,\dots,x_n)$ be an n-tuple of real numbers and let $\boldsymbol{p} = (p_1,\dots,p_n)$ be an n-tuple of positive numbers. Then $A(\boldsymbol{x}) = \frac{1}{P_n}\sum_{i=1}^n p_i x_i$ is a normalized positive linear functional, so from Theorem 8.4, inequality (8.8), we have

$$\begin{aligned}
0 &\le \sum_{i=1}^n p_i f(x_i) - f\left(\sum_{i=1}^n p_i x_i\right) \\
&\le (M-\overline{x})(\overline{x}-m) \sup_{t\in\langle m,M\rangle} \Psi_f(t;m,M) - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^n p_i \left|x_i - \frac{m+M}{2}\right|\right)\delta_f \\
&\le \frac{(M-\overline{x})(\overline{x}-m)}{M-m}(f'_-(M) - f'_+(m)) - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^n p_i \left|x_i - \frac{m+M}{2}\right|\right)\delta_f \\
&\le \frac{1}{4}(M-m)(f'_-(M) - f'_+(m)) - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^n p_i \left|x_i - \frac{m+M}{2}\right|\right)\delta_f,
\end{aligned} \tag{8.37}$$

Since $\boldsymbol{q} = (q_1,\dots,q_n)$ are nonnegative real numbers, we can put

$$p_i = \frac{q_i}{\sum_{i=1}^n q_i} = \frac{q_i}{Q_n} \quad \text{and} \quad x_i = \frac{p_i}{q_i}$$

in (8.37) and get

$$\begin{aligned}
0 &\le \sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i} f\left(\frac{p_i}{q_i}\right) - f\left(\sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i}\frac{p_i}{q_i}\right) \\
&\le \left(M - \sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i}\frac{p_i}{q_i}\right)\left(\sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i}\frac{p_i}{q_i} - m\right) \sup_{t\in\langle m,M\rangle} \Psi_f(t;m,M) \\
&\quad - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i}\left|\frac{p_i}{q_i} - \frac{m+M}{2}\right|\right)\delta_f \\
&\le \frac{f'_-(M) - f'_+(m)}{M-m}\left(M - \sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i}\frac{p_i}{q_i}\right)\left(\sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i}\frac{p_i}{q_i} - m\right) \\
&\quad - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i}\left|\frac{p_i}{q_i} - \frac{m+M}{2}\right|\right)\delta_f \\
&\le \frac{1}{4}(M-m)(f'_-(M) - f'_+(m)) - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^n \frac{q_i}{\sum_{i=1}^n q_i}\left|\frac{p_i}{q_i} - \frac{m+M}{2}\right|\right)\delta_f,
\end{aligned}$$

and after multiplying by Q_n we get (8.36). □

The result that follows is a special cases of Theorem 8.18. It gives us different bounds of those that we have already obtained for the Kullback-Leibler divergence of two probability distributions.

Corollary 8.2 ([25]) *Let $[m,M] \subset \mathbb{R}$ be an interval and let us assume that the base of the logarithm is greater than* 1.

- *Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{q} = (q_1,\ldots,q_n)$ be n-tuples of nonnegative real numbers such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,n$. Then*

$$
\begin{aligned}
0 &\le \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} - P_n \log\left(\frac{P_n}{Q_n}\right) \\
&\le Q_n \left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right) \sup_{t\in\langle m,M\rangle} \Psi_{id\cdot\log}(t;m,M) \\
&\quad - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right) \\
&\le \frac{Q_n}{M-m}\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right)\log\frac{M}{m} \qquad (8.38)\\
&\quad - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right) \\
&\le \frac{Q_n}{4}(M-m)\log\frac{M}{m} \\
&\quad - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right).
\end{aligned}
$$

- *Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{q} = (q_1,\ldots,q_n) \in \mathbb{P}$ be probability distributions such that $m \le p_i/q_i \le M$ holds for every $i = 1,\ldots,n$. Then*

$$
\begin{aligned}
0 &\le D_{KL}(\boldsymbol{p},\boldsymbol{q}) \\
&\le (M-1)(1-m) \sup_{t\in\langle m,M\rangle} \Psi_{id\cdot\log}(t;m,M) \\
&\quad - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right) \\
&\le \frac{1}{M-m}(M-1)(1-m)\log\frac{M}{m} \qquad (8.39)\\
&\quad - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right) \\
&\le \frac{1}{4}(M-m)\log\frac{M}{m} \\
&\quad - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right).
\end{aligned}
$$

If the base of the logarithm is less than 1*, the inequality sign in the inequalities above is reversed.*

Proof. Let $\boldsymbol{p} = (p_1, \ldots, p_n)$ and $\boldsymbol{q} = (q_1, \ldots, q_n)$ be an n-tuples of nonnegative real numbers. Function $t \mapsto t\log t$ is convex, so inequality (8.38) follows from Theorem 8.18, inequality (8.36), by setting $f(t) = t\log t$.

Inequality (8.39) is a special case of the inequality (8.38) for probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$. □

Remark 8.4 If in Theorem 8.18, inequality (8.36), we set $f(t) = -\log t$ with the base greater than 1, we get the following:

- for n-tuples of nonnegative real numbers $\boldsymbol{p} = (p_1, \ldots, p_n)$ and $\boldsymbol{q} = (q_1, \ldots, q_n)$ such that $p_i/q_i \in [m, M]$ for every $i = 1, \ldots, n$ we have

$$
\begin{aligned}
0 &\leq \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right) + Q_n \log\left(\frac{P_n}{Q_n}\right) \\
&\leq Q_n\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right) \sup_{t\in\langle m,M\rangle} \Psi_{-\log}(t; m, M) \\
&\quad - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\log\frac{(m+M)^2}{4mM} \qquad (8.40)\\
&\leq \frac{Q_n}{Mm}\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right) \\
&\quad - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\log\frac{(m+M)^2}{4mM} \\
&\leq \frac{Q_n(M-m)^2}{4Mm} - \left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\log\frac{(m+M)^2}{4mM}.
\end{aligned}
$$

- for probability distributions $\boldsymbol{p} = (p_1, \ldots, p_n)$ and $\boldsymbol{q} = (q_1, \ldots, q_n) \in \mathbb{P}$ such that $m \leq p_i/q_i \leq M$ holds for every $i = 1, \ldots, n$ we have

$$
\begin{aligned}
0 &\leq D_{KL}(\boldsymbol{q}, \boldsymbol{p}) \\
&\leq (M-1)(1-m) \sup_{t\in\langle m,M\rangle} \Psi_{-\log}(t; m, M) \\
&\quad - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\log\frac{(m+M)^2}{4mM} \qquad (8.41)\\
&\leq \frac{1}{Mm}(M-1)(1-m) \\
&\quad - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\log\frac{(m+M)^2}{4mM} \\
&\leq \frac{(M-m)^2}{4Mm} - \left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\log\frac{(m+M)^2}{4mM}.
\end{aligned}
$$

If the base of the logarithm is less than 1, the inequality sign in the inequalities above is reversed.

By following the same steps as in the proof of Theorem 8.18, but starting from Theorem 8.5, we get lower and upper bounds for the difference in the results from Theorem 8.17, and consequently in Dragomir's result (8.2).

Theorem 8.19 ([25]) *Let $I \subset \mathbb{R}$ be an interval such that its interior contains the interval $[m,M]$, let $f\colon I \to \mathbb{R}$ be a continuous function and let δ_f be defined in (8.7). Let $\boldsymbol{p} = (p_1,\dots,p_n)$ be an n-tuple of real numbers and $\boldsymbol{q} = (q_1,\dots,q_n)$ be an n-tuple of nonnegative real numbers such that $p_i/q_i \in [m,M]$ for every $i = 1,\dots,n$. Let Ψ_f be defined in (8.10). If the function f is convex, then we have*

$$\begin{aligned}
&\left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\delta_f \\
\leq &\frac{MQ_n - P_n}{M-m}f(m) + \frac{P_n - mQ_n}{M-m}f(M) - \hat{D}_f(P,Q) \\
\leq &\sup_{t\in\langle m,M\rangle}\Psi_f(t;m,M)\sum_{i=1}^{n}\left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right) \\
\leq &Q_n\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right)\sup_{t\in\langle m,M\rangle}\Psi_f(t;m,M) \\
\leq &Q_n\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right)\frac{f'_-(M) - f'_+(m)}{M-m} \\
\leq &\frac{Q_n}{4}(M-m)(f'_-(M) - f'_+(m)).
\end{aligned} \tag{8.42}$$

If the function f is concave, the inequality signs are reversed.

We can utilize Theorem 8.19 to obtain lower and upper bounds for the difference in the results from Corollary 8.1, as well as for the reversed Kullback-Leibler divergence.

Corollary 8.3 ([25]) *Let $[m,M] \subset \mathbb{R}$ be an interval and let us assume that the base of the logarithm is greater than* 1.

- *Let $\boldsymbol{p} = (p_1,\dots,p_n)$ and $\boldsymbol{q} = (q_1,\dots,q_n)$ be n-tuples of nonnegative real numbers such that $p_i/q_i \in [m,M]$ for every $i = 1,\dots,n$. Then*

$$\begin{aligned}
&\left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right) \\
\leq &Q_n\frac{Mm}{M-m}\log\left(\frac{m}{M}\right) + \frac{P_n}{M-m}\log\left(\frac{M^M}{m^m}\right) - \sum_{i=1}^{n}p_i\log\left(\frac{p_i}{q_i}\right) \\
\leq &\sup_{t\in\langle m,M\rangle}\Psi_{id\cdot\log}(t;m,M)\sum_{i=1}^{n}\left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq Q_n \left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right) \sup_{t \in \langle m,M \rangle} \Psi_{id \cdot \log}(t;m,M) \qquad (8.43)\\
&\leq \frac{Q_n}{M-m}\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right)\log\left(\frac{M}{m}\right) \leq \frac{Q_n}{4}(M-m)\log\left(\frac{M}{m}\right).
\end{aligned}
$$

- *Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{q} = (q_1,\ldots,q_n) \in \mathbb{P}$ be probability distributions such that $m \leq p_i/q_i \leq M$ holds for every $i = 1,\ldots,n$. Then*

$$
\begin{aligned}
&\left(\frac{1}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\left(m\log\frac{2m}{m+M} + M\log\frac{2M}{m+M}\right)\\
&\leq \frac{Mm}{M-m}\log\left(\frac{m}{M}\right) + \frac{1}{M-m}\log\left(\frac{M^M}{m^m}\right) - D_{KL}(\boldsymbol{p},\boldsymbol{q})\\
&\leq \sup_{t \in \langle m,M \rangle} \Psi_{id \cdot \log}(t;m,M)\sum_{i=1}^{n}\left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right)\\
&\leq (M-1)(1-m) \sup_{t \in \langle m,M \rangle} \Psi_{id \cdot \log}(t;m,M) \qquad (8.44)\\
&\leq \frac{1}{M-m}(M-1)(1-m)\log\left(\frac{M}{m}\right) \leq \frac{1}{4}(M-m)\log\left(\frac{M}{m}\right).
\end{aligned}
$$

If the base of the logarithm is less than 1*, the inequality sign in the inequalities above is reversed.*

Remark 8.5 As in Remark 8.4, we can set $f(t) = -\log t$ with the base greater than 1 in Theorem 8.19, inequality (8.42), and obtain the following inequalities for the reversed Kullback-Leibler divergence:

- for n-tuples of nonnegative real numbers $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{q} = (q_1,\ldots,q_n)$ such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,n$ we have

$$
\begin{aligned}
&\left(\frac{Q_n}{2} - \frac{1}{M-m}\sum_{i=1}^{n}\left|p_i - \frac{m+M}{2}q_i\right|\right)\log\frac{(m+M)^2}{4mM}\\
&\leq \frac{Q_n}{M-m}\log\left(\frac{M^m}{m^M}\right) + \frac{P_n}{M-m}\log\left(\frac{m}{M}\right) - \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right)\\
&\leq \sup_{t \in \langle m,M \rangle} \Psi_{\log}(t;m,M)\sum_{i=1}^{n}\left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right)\\
&\leq Q_n \left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right) \sup_{t \in \langle m,M \rangle} \Psi_{\log}(t;m,M)\\
&\leq -\frac{Q_n}{Mm}\left(M - \frac{P_n}{Q_n}\right)\left(\frac{P_n}{Q_n} - m\right) \leq -\frac{Q_n}{4Mm}(M-m)^2. \qquad (8.45)
\end{aligned}
$$

- for probability distributions $\boldsymbol{p} = (p_1, \dots, p_n)$ and $\boldsymbol{q} = (q_1, \dots, q_n) \in \mathbb{P}$ such that $m \leq p_i/q_i \leq M$ holds for every $i = 1, \dots, n$ we have

$$\begin{aligned}
&\left(\frac{1}{2} - \frac{1}{M-m} \sum_{i=1}^{n} \left| p_i - \frac{m+M}{2} q_i \right| \right) \log \frac{(m+M)^2}{4mM} \\
&\leq \frac{1}{M-m} \log \left(\frac{M^{m-1}}{m^{M-1}} \right) - D_{KL}(\boldsymbol{q}, \boldsymbol{p}) \\
&\leq \sup_{t \in \langle m, M \rangle} \Psi_{\log}(t; m, M) \sum_{i=1}^{n} \left(M - \frac{p_i}{q_i} \right) \left(\frac{p_i}{q_i} - m \right) \\
&\leq (M-1)(1-m) \sup_{t \in \langle m, M \rangle} \Psi_{\log}(t; m, M) \\
&\leq -\frac{1}{Mm}(M-1)(1-m) \leq -\frac{1}{4Mm}(M-m)^2. \qquad (8.46)
\end{aligned}$$

If the base of the logarithm is less than 1, the inequality sign in the inequalities above is reversed.

Unlike previous results, the following results do not require convexity in the classical sense of the function f. We start with an Edmundson-Lah-Ribarič type inequality for the generalized f-divergence functional $\tilde{D}_f(\boldsymbol{p}, \boldsymbol{q})$, where the function f has bounded second order divided differences. This is a significant progress in relation to the previous results, since the class of functions with bounded second order divided differences is much greater then the class of convex functions.

Theorem 8.20 ([26]) *Let $[m, M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$ and let $f \colon [m, M] \to \mathbb{R}$ be a function with $\gamma \leq [m, t, M]f \leq \Gamma$. Let $\boldsymbol{p} = (p_1, \dots, p_n)$ and $\boldsymbol{p} = (q_1, \dots, q_n)$ be probability distributions such that $p_i/q_i \in [m, M]$ for every $i = 1, \dots, n$. Then we have*

$$\begin{aligned}
\gamma \sum_{i=1}^{n} q_i \left(M - \frac{p_i}{q_i} \right) \left(\frac{p_i}{q_i} - m \right) &\leq \frac{M-1}{M-m} f(m) + \frac{1-m}{M-m} f(M) - \tilde{D}_f(\boldsymbol{p}, \boldsymbol{q}) \\
&\leq \Gamma \sum_{i=1}^{n} q_i \left(M - \frac{p_i}{q_i} \right) \left(\frac{p_i}{q_i} - m \right). \qquad (8.47)
\end{aligned}$$

Proof. The function f has bounded second order divided difference with bounds γ and Γ, so when we set linear functional A from (8.14) to be a discrete sum, we get

$$\begin{aligned}
\gamma \sum_{i=1}^{n} p_i (M - x_i)(x_i - m) &\leq \frac{M - \bar{x}}{M-m} \phi(m) + \frac{\bar{x} - m}{M-m} \phi(M) - \sum_{i=1}^{n} p_i \phi(x_i) \\
&\leq \Gamma \sum_{i=1}^{n} p_i (M - x_i)(x_i - m), \qquad (8.48)
\end{aligned}$$

where $\boldsymbol{x} = (x_1, \dots, x_n)$ is an n-tuple of real numbers from $[m, M]$, $\boldsymbol{p} = (p_1, \dots, p_n)$ is an n-tuple of nonnegative real numbers such that $\sum_{i=1}^{n} p_i = 1$, and $\bar{x} = \sum_{i=1}^{n} p_i x_i$. Now, in the relation (8.48) we can put

$$\phi = f, \quad p_i = q_i \text{ and } x_i = \frac{p_i}{q_i},$$

and after calculating

$$\bar{x} = \sum_{i=1}^{n} q_i \frac{p_i}{q_i} = \sum_{i=1}^{n} p_i = 1$$

we get (8.47). □

By following the same idea as in the proof of the previous theorem, but starting with the relation (8.17) from Theorem 8.7, we get the following result, which is a Jensen type inequality for the generalized f-divergence functional $\tilde{D}_f(\boldsymbol{p},\boldsymbol{q})$.

Theorem 8.21 *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \le 1 \le M$ and let $f\colon [m,M] \to \mathbb{R}$ be a function with $\gamma \le [m,t,M]f \le \Gamma$. Let $\boldsymbol{p} = (p_1,\dots,p_n)$ and $\boldsymbol{p} = (q_1,\dots,q_n)$ be probability distributions such that $p_i/q_i \in [m,M]$ for every $i = 1,\dots,n$. Then we have*

$$\begin{aligned}
&\gamma(M-1)(1-m) - \Gamma \sum_{i=1}^{n} q_i \left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right) \qquad (8.49)\\
\le&\tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) - f(1) \le \Gamma(M-1)(1-m) - \gamma \sum_{i=1}^{n} q_i \left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right).
\end{aligned}$$

Remark 8.6 If the function $f\colon [m,M] \to \mathbb{R}$ is additionally convex, then from (8.15), by following the same idea as in the proof of Theorem 8.20, we get Edmundson-Lah-Ribarič type inequality for the Csiszár f-divergence functional:

$$\begin{aligned}
0 \le&\gamma \sum_{i=1}^{n} q_i \left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right)\\
\le&\frac{M-1}{M-m} f(m) + \frac{1-m}{M-m} f(M) - D_f(\boldsymbol{p},\boldsymbol{q})\\
\le&\Gamma \sum_{i=1}^{n} q_i \left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right) \qquad (8.50)\\
\le&\Gamma(M-1)(1-m) \le \frac{\Gamma}{4}(M-m)^2.
\end{aligned}$$

Jensen type inequality for Csiszár divergence functional is a special case of Theorem 8.21 for a convex function.

The generating function of the Kullback-Leibler divergence $f(t) = t\log t$ is convex, and its second order divided difference $[m,t,M]f$ is a continuous and decreasing function, which means that it attains its maximal and minimal value in the points m and M respectively.

We calculate the bounds for the second order divided difference of the function $f(t) = t\log t$:

$$\Gamma = [m,m,M]\mathrm{id}\cdot\log = \frac{1}{M-m}\left(\frac{M\log M - m\log m}{M-m} - [(\mathrm{id}\cdot\log)(m)]'_{+}\right)$$

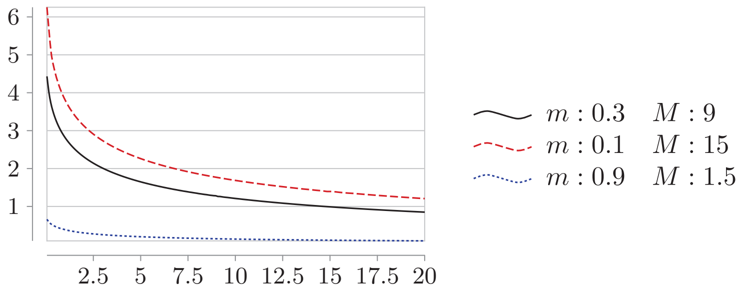

Figure 8.1: Graphs of the Function $-[m,t,M]\mathrm{id}\circ\log$ for different Choices of the Points m and M.

$$
\begin{aligned}
&= \frac{1}{M-m}\left(\frac{1}{M-m}\log\frac{M^M}{m^m} - \log m - 1\right)\\
\gamma &= [m,M,M]\mathrm{id}\cdot\log\frac{1}{M-m}\left([(\mathrm{id}\cdot\log)(m)]'_- - \frac{M\log M - m\log m}{M-m}\right)\\
&= \frac{1}{M-m}\left(\log M + 1 - \frac{1}{M-m}\log\frac{M^M}{m^m}\right).
\end{aligned}
$$

Now, as a special case of Theorem 8.20 and Theorem 8.21 for $f(t) = t\log t$, taking into account convexity of the function f, we have obtained Jensen and Edmundson-Lah-Ribarič type inequalities for Kullback-Leibler divergence $D_{KL}(\boldsymbol{p},\boldsymbol{q})$.

Corollary 8.4 ([26]) *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$. Let $\boldsymbol{p} = (p_1,\dots,p_n)$ and $\boldsymbol{p} = (q_1,\dots,q_n)$ be probability distributions such that $p_i/q_i \in [m,M]$ for every $i = 1,\dots,n$. Then we have*

$$
\begin{aligned}
0 &\leq \frac{1}{M-m}\left(\log M + 1 - \frac{1}{M-m}\log\frac{M^M}{m^m}\right)\sum_{i=1}^n q_i\left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right)\\
&\leq \frac{mM}{M-m}\log\frac{m}{M} + \frac{1}{M-m}\log\frac{M^M}{m^m} - D_{KL}(\boldsymbol{p},\boldsymbol{q}) \qquad (8.51)\\
&\leq \frac{1}{M-m}\left(\frac{1}{M-m}\log\frac{M^M}{m^m} - \log m - 1\right)\sum_{i=1}^n q_i\left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right)\\
&\leq \frac{1}{M-m}\left(\frac{1}{M-m}\log\frac{M^M}{m^m} - \log m - 1\right)(M-1)(1-m)\\
&\leq \frac{1}{4}\left(\log\frac{M^M}{m^m} - (\log m + 1)(M-m)\right).
\end{aligned}
$$

Corollary 8.5 ([26]) *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$. Let $\boldsymbol{p} = (p_1,\dots,p_n)$ and $\boldsymbol{p} = (q_1,\dots,q_n)$ be probability distributions such that $p_i/q_i \in [m,M]$ for* *every $i = 1,\ldots,n$. Then we have*

$$0 \leq D_{KL}(\boldsymbol{p},\boldsymbol{q}) \leq \frac{1}{M-m}\left[\left(\frac{1}{M-m}\log\frac{M^M}{m^m} - \log m - 1\right)(M-1)(1-m)\right.$$
$$\left. - \left(\log M + 1 - \frac{1}{M-m}\log\frac{M^M}{m^m}\right)\sum_{i=1}^{n} q_i\left(M - \frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i} - m\right)\right]. \tag{8.52}$$

The function $f(t) = -\log t$ is also convex, and its second order divided difference $[m,t,M]f$ is a continuous and decreasing function, which means that it attains its maximal and minimal value in the points m and M respectively.

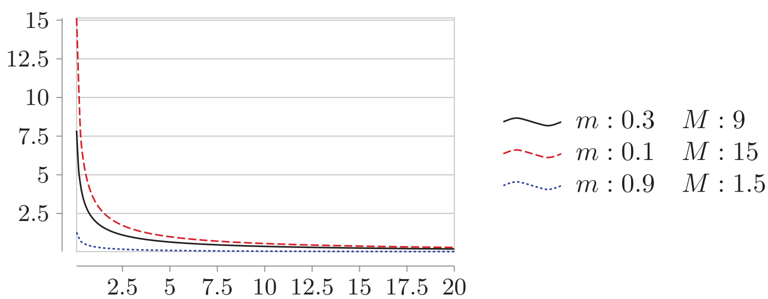

Figure 8.2: Graphs of the function $-[m,t,M]\log$ for different choices of the points m and M.

We calculate the bounds for the second order divided difference of the function $f(t) = -\log t$:

$$\begin{aligned}\Gamma = -[m,m,M]\log &= \frac{1}{M-m}\left(\frac{-\log M + \log m}{M-m} - (-\log)'_+(m)\right)\\ &= \frac{1}{M-m}\left(\frac{1}{M-m}\log\frac{m}{M} + \frac{1}{m}\right)\\ \gamma = -[m,M,M]\log &= \frac{1}{M-m}\left((-\log)'_-(M) - \frac{-\log M + \log m}{M-m}\right)\\ &= -\frac{1}{M-m}\left(\frac{1}{M} + \frac{1}{M-m}\log\frac{m}{M}\right).\end{aligned}$$

As a special case of Theorem 8.20 and Theorem 8.21 for $f(t) = -\log t$, taking into account convexity of the function f, we get Jensen and Edmundson-Lah-Ribarič type inequalities for the reversed Kullback-Leibler divergence $D_{KL}(\boldsymbol{q},\boldsymbol{p})$.

Corollary 8.6 ([26]) *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$. Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{p} = (q_1,\ldots,q_n)$ be probability distributions such that $p_i/q_i \in [m,M]$ for* *every $i = 1,\ldots,n$. Then we have*

$$
\begin{aligned}
0 \leq & -\frac{1}{M-m}\left(\frac{1}{M}+\frac{1}{M-m}\log\frac{m}{M}\right)\sum_{i=1}^{n} q_i\left(M-\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-m\right) \\
\leq & \frac{1}{M-m}\log\frac{M^m}{m^M}+\frac{1}{M-m}\log\frac{m}{M}-D_{KL}(\boldsymbol{q},\boldsymbol{p}) \qquad (8.53) \\
\leq & \frac{1}{M-m}\left(\frac{1}{M-m}\log\frac{m}{M}+\frac{1}{m}\right)\sum_{i=1}^{n} q_i\left(M-\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-m\right) \\
\leq & \frac{1}{M-m}\left(\frac{1}{M-m}\log\frac{m}{M}+\frac{1}{m}\right)(M-1)(1-m) \\
\leq & \frac{1}{4}\left(\log\frac{m}{M}+\frac{M}{m}-1\right).
\end{aligned}
$$

Corollary 8.7 ([26]) *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$. Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{p} = (q_1,\ldots,q_n)$ be probability distributions such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,n$. Then we have*

$$
\begin{aligned}
0 \leq D_{KL}(\boldsymbol{q},\boldsymbol{p}) \leq & \frac{1}{M-m}\Bigg[\left(\frac{1}{M-m}\log\frac{m}{M}+\frac{1}{m}\right)(M-1)(1-m) \\
& +\left(\frac{1}{M}+\frac{1}{M-m}\log\frac{m}{M}\right)\sum_{i=1}^{n} q_i\left(M-\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-m\right)\Bigg]. \qquad (8.54)
\end{aligned}
$$

Following results are applications of Theorem 8.9 and Theorem 8.10, and they provide us with an Edmundson-Lah-Ribarič type and Jensen type inequality respectively for the generalized f-divergence functional for 3-convex function.

Theorem 8.22 ([24]) *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$ and let $f\colon [m,M] \to \mathbb{R}$ be a 3-convex function. Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{p} = (q_1,\ldots,q_n)$ be probability distributions such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,n$. Then we have*

$$
\begin{aligned}
& \frac{1}{M-m}\sum_{i=1}^{n} q_i\left(M-\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-m\right)\left(\frac{f(M)-f(m)}{M-m}-f'_+(m)\right) \\
\leq & \frac{M-1}{M-m}f(m)+\frac{1-m}{M-m}f(M)-\tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \qquad (8.55) \\
\leq & \frac{1}{M-m}\sum_{i=1}^{n} q_i\left(M-\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-m\right)\left(f'_-(M)-\frac{f(M)-f(m)}{M-m}\right)
\end{aligned}
$$

Proof. Let $\boldsymbol{x} = (x_1,\ldots,x_n)$ such that $x_i \in [m,M]$ for $i = 1,\ldots,n$. For a 3-convex function ϕ, in the relation (8.22) we can replace

$$f \longleftrightarrow \boldsymbol{x}, \quad \text{and} \quad A(\boldsymbol{x}) = \sum_{i=1}^{n} p_i x_i.$$

In that way we get

$$\begin{aligned}
&\frac{\sum_{i=1}^{n} p_i(M-x_i)(x_i-m)}{M-m}\left(\frac{\phi(M)-\phi(m)}{M-m}-\phi'_+(m)\right)\\
\le&\frac{M-\bar{x}}{M-m}\phi(m)+\frac{\bar{x}-m}{M-m}\phi(M)-\sum_{i=1}^{n} p_i\phi(x_i)\\
\le&\frac{\sum_{i=1}^{n} p_i(M-x_i)(x_i-m)}{M-m}\left(\phi'_-(M)-\frac{\phi(M)-\phi(m)}{M-m}\right),
\end{aligned}$$

where $\bar{x}=\sum_{i=1}^{n} p_i x_i$. Since the function f is 3-convex, in the previous relation we can set

$$\phi=f,\ \ p_i=q_i \ \text{ and } \ x_i=\frac{p_i}{q_i},$$

and after calculating

$$\bar{x}=\sum_{i=1}^{n} q_i\frac{p_i}{q_i}=\sum_{i=1}^{n} p_i=1$$

we get (8.55). □

Theorem 8.23 ([24]) *Let $[m,M]\subset\mathbb{R}$ be an interval such that $m\le 1\le M$ and let $f\colon [m,M]\to\mathbb{R}$ be a 3-convex function. Let $\boldsymbol{p}=(p_1,\dots,p_n)$ and $\boldsymbol{p}=(q_1,\dots,q_n)$ be probability distributions such that $p_i/q_i\in[m,M]$ for every $i=1,\dots,n$. Then we have*

$$\begin{aligned}
&\frac{(M-1)(1-m)}{M-m}\left(\frac{f(M)-f(m)}{M-m}-f'_+(m)\right)\\
&-\frac{1}{M-m}\sum_{i=1}^{n} q_i\left(M-\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-m\right)\left(f'_-(M)-\frac{f(M)-f(m)}{M-m}\right)\\
\le&\tilde{D}_f(\boldsymbol{p},\boldsymbol{q})-f(1)\le\frac{(M-1)(1-m)}{M-m}\left(f'_-(M)-\frac{f(M)-f(m)}{M-m}\right)\\
&-\frac{1}{M-m}\sum_{i=1}^{n} q_i\left(M-\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-m\right)\left(\frac{f(M)-f(m)}{M-m}-f'_+(m)\right).
\end{aligned}\tag{8.56}$$

Proof. As in the proof of the previous theorem, let $\boldsymbol{x}=(x_1,\dots,x_n)$ such that $x_i\in[m,M]$ for $i=1,\dots,n$. For a 3-convex function ϕ, in the relation (8.23) we can replace

$$f\longleftrightarrow\boldsymbol{x},\ \ \text{and}\ \ A(\boldsymbol{x})=\sum_{i=1}^{n} p_i x_i$$

and obtain the following discrete sequence of inequalities:

$$\begin{aligned}
&\frac{(M-\bar{x})(\bar{x}-m)}{M-m}\left(\frac{\phi(M)-\phi(m)}{M-m}-\phi'_+(m)\right)\\
&-\frac{\sum_{i=1}^{n} p_i(M-x_i)(x_i-m)}{M-m}\left(\phi'_-(M)-\frac{\phi(M)-\phi(m)}{M-m}\right)
\end{aligned}$$

$$\le \sum_{i=1}^{n} p_i\phi(x_i) - \phi(\bar{x}) \le \frac{(M-\bar{x})(\bar{x}-m)}{M-m}\left(\phi'_-(M) - \frac{\phi(M)-\phi(m)}{M-m}\right)$$
$$- \frac{\sum_{i=1}^{n} p_i(M-x_i)(x_i-m)}{M-m}\left(\frac{\phi(M)-\phi(m)}{M-m} - \phi'_+(m)\right).$$

The function f is 3-convex, so in the previous relation we can set

$$\phi = f, \quad p_i = q_i \text{ and } x_i = \frac{p_i}{q_i},$$

and after calculating

$$\bar{x} = \sum_{i=1}^{n} q_i \frac{p_i}{q_i} = \sum_{i=1}^{n} p_i = 1$$

we get (8.56). □

Next two results are obtained as an application of Theorem 8.11 and Theorem 8.12 respectively, and they give us different Edmundson-Lah-Ribarič type inequalities for the generalized f-divergence functional.

Theorem 8.24 ([27]) *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \le 1 \le M$. Let f be a 3-convex function on the interval I whose interior contains $[m,M]$ and differentiable on $\langle m,M\rangle$. Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{p} = (q_1,\ldots,q_n)$ be probability distributions such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,n$. Then we have*

$$\begin{aligned}
&(1-m)\left[\frac{f(M)-f(m)}{M-m} - \frac{f'_+(m)}{2}\right] - \frac{1}{2}\sum_{i=1}^{n}(p_i - mq_i) f'\left(\frac{p_i}{q_i}\right)\\
&\le \frac{M-1}{M-m} f(m) + \frac{1-m}{M-m} f(M) - \tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \qquad (8.57)\\
&\le \frac{1}{2}\sum_{i=1}^{n}(Mq_i - p_i) f'\left(\frac{p_i}{q_i}\right) - (M-1)\left[\frac{f(M)-f(m)}{M-m} - \frac{f'_-(M)}{2}\right].
\end{aligned}$$

Proof. Let $\boldsymbol{x} = (x_1,\ldots,x_n)$ such that $x_i \in [m,M]$ for $i = 1,\ldots,n$. Let ϕ be a 3-convex function on the interval I whose interior contains $[m,M]$ and differentiable on $\langle m,M\rangle$. In the relation (8.24) we can replace

$$f \longleftrightarrow \boldsymbol{x}, \quad \text{and } A(\boldsymbol{x}) = \sum_{i=1}^{n} p_i x_i.$$

In that way we get

$$\begin{aligned}
&(\bar{x}-m)\left[\frac{\phi(M)-\phi(m)}{M-m} - \frac{\phi'_+(m)}{2}\right] - \frac{1}{2}\sum_{i=1}^{n} p_i(x_i - m)\phi'(x_i)\\
&\le \frac{M-\bar{x}}{M-m}\phi(m) + \frac{\bar{x}-m}{M-m}\phi(M) - \sum_{i=1}^{n} p_i\phi(x_i)\\
&\le \frac{1}{2}\sum_{i=1}^{n} p_i(M-x_i)\phi'(x_i) - (M-\bar{x})\left[\frac{\phi(M)-\phi(m)}{M-m} - \frac{\phi'_-(M)}{2}\right]
\end{aligned}$$

where $\bar{x} = \sum_{i=1}^{n} p_i x_i$. Since the function f satisfies the same assumtions as ϕ, in the previous relation we can set

$$\phi = f, \quad p_i = q_i \text{ and } x_i = \frac{p_i}{q_i},$$

and after calculating

$$\bar{x} = \sum_{i=1}^{n} q_i \frac{p_i}{q_i} = \sum_{i=1}^{n} p_i = 1$$

we get (8.57). □

By utilizing Theorem 8.12 in the analogous way as above, we get a different Edmundson-Lah-Ribarič type inequality for the generalized f-divergence functional, and it is given in the following theorem.

Theorem 8.25 ([27]) *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \le 1 \le M$. Let f be a 3-convex function on the interval I whose interior contains $[m,M]$ and differentiable on $\langle m,M \rangle$. Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{p} = (q_1,\ldots,q_n)$ be probability distributions such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,n$. Then we have*

$$\begin{aligned}
&(M-1)\left[f'_{-}(M) - \frac{f(M)-f(m)}{M-m}\right] - \frac{f''_{-}(M)}{2}\sum_{i=1}^{n}\frac{(Mq_i-p_i)^2}{q_i} \\
\le &\frac{M-1}{M-m}f(m) + \frac{1-m}{M-m}f(M) - \tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \qquad (8.58)\\
\le &(1-m)\left[\frac{f(M)-f(m)}{M-m} - f'_{+}(m)\right] - \frac{f''_{+}(m)}{2}\sum_{i=1}^{n}\frac{(p_i-mq_i)^2}{q_i}.
\end{aligned}$$

Remark 8.7 Let $\boldsymbol{p} = (p_1,\ldots,p_n)$ and $\boldsymbol{p} = (q_1,\ldots,q_n)$ be probability distributions and let $[m,M] \subset \mathbb{R}$ be an interval such that $m \le 1 \le M$ and $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,n$.

▷ **Kullback-Leibler divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_{KL}(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n} q_i \log \frac{q_i}{p_i},$$

and the corresponding generating function $f(t) = t\log t, t > 0$. We can calculate $f'''(t) = -\frac{1}{t^2} < 0$, so the function $-f(t) = -t\log t$ is 3-convex. It is obvious that for the Kullback-Leibler divergence the inequalities (8.55), (8.56), (8.57) and (8.58) hold with reversed signs of inequality, with

$$f'_{+}(m) = \log m + 1, \quad f'_{-}(M) = \log M + 1$$

and

$$f''_{+}(m) = \frac{1}{m}, \quad f''_{-}(M) = \frac{1}{M}.$$

▷ **Hellinger divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

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

with the corresponding generating function $f(t) = \frac{1}{2}(1-\sqrt{t})^2, t > 0$. We see that $f'''(t) = -\frac{3}{8}t^{-\frac{5}{2}} < 0$, so the function $-f(t) = -\frac{1}{2}(1-\sqrt{t})^2$ is 3-convex. For the Hellinger divergence the inequalities (8.55), (8.56), (8.57) and (8.58) hold with reversed signs of inequality, with

$$f'_+(m) = -\frac{1}{2\sqrt{m}} + \frac{1}{2}, \;\; f'_-(M) = -\frac{1}{2\sqrt{M}} + \frac{1}{2}$$

and

$$f''_+(m) = \frac{1}{4\sqrt{m^3}}, \;\; f''_-(M) = \frac{1}{4\sqrt{M^3}}.$$

▷ **Renyi divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_\alpha(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n} q_i^{\alpha-1} p_i^{\alpha}, \;\; \alpha \in \mathbb{R},$$

and the corresponding generating function is $f(t) = t^\alpha, t > 0$. We calculate that $f'''(t) = \alpha(\alpha-1)(\alpha-2)t^{\alpha-3}$, which is 3-convex for $0 \le \alpha \le 1$ and $\alpha \ge 2$, and $-f(t) = -t^\alpha$ is 3-convex for $\alpha \le 0$ and $1 < \alpha < 2$. We have

$$f'_+(m) = \alpha m^{\alpha-1}, \;\; f'_-(M) = \alpha M^{\alpha-1},$$

$$f''_+(m) = \alpha(\alpha-1)m^{\alpha-2} \;\text{ and }\; f''_-(M) = \alpha(\alpha-1)M^{\alpha-2}.$$

As regards the Renyi divergence, the inequalities (8.55), (8.56), (8.57) and (8.58) hold for $0 \le \alpha \le 1$ and $\alpha \ge 2$, and if $\alpha \le 0$ or $1 < \alpha < 2$ the signs of inequality are reversed.

▷ **Harmonic divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_{Ha}(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n} \frac{2p_i q_i}{p_i + q_i},$$

and the corresponding generating function $f(t) = \frac{2t}{1+t}$. We can calculate $f'''(t) = \frac{12}{(1+t)^4} > 0$, so the function f is 3-convex. It is obvious that for the harmonic divergence the inequalities (8.55), (8.56), (8.57) and (8.58) hold with

$$f'_+(m) = \frac{2}{(1+m)^2}, \;\; f'_-(M) = \frac{2}{(1+M)^2}$$

and

$$f''_+(m) = -\frac{4}{(1+m)^3}, \;\; f''_-(M) = -\frac{4}{(1+M)^3}.$$

▷ **Jeffreys divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_J(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{2}\sum_{i=1}^{n}(q_i - p_i)\log\frac{q_i}{p_i},$$

with the corresponding generating function $f(t) = (1-t)\log\frac{1}{t}, t > 0$. We see that $f'''(t) = -\frac{1}{t^2} - \frac{2}{t^3} < 0$, so the function $-f(t) = (1-t)\log t$ is 3-convex, and we instantly get that for the Jeffreys divergence the inequalities (8.55), (8.56), (8.57) and (8.58) hold with reversed signs of inequality, with

$$f'_+(m) = \log m - \frac{1}{m} + 1, \quad f'_-(M) = \log M - \frac{1}{M} + 1$$

and

$$f''_+(m) = \frac{1}{m} + \frac{1}{m^2}, \quad f''_-(M) = \frac{1}{M} + \frac{1}{M^2}.$$

The results that follow are a generalization of the previous results which hold for the class of n-convex functions. Until the end of this section, when mentioning the interval $[m,M]$, we assume that $[m,M] \subseteq \mathbb{R}_+$.

We can utilize Theorem 8.13 to get an Edmundson-Lah-Ribarič type inequality for the above defined generalized f-divergence functional.

Theorem 8.26 *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$. Let $f \in \mathscr{C}^n([m,M])$ and let $\boldsymbol{p} = (p_1,\ldots,p_r)$ and $\boldsymbol{p} = (q_1,\ldots,q_r)$ be probability distributions such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,r$. If the function f is n-convex and if n and $3 \leq v \leq n-1$ are of different parity, then*

$$\frac{M-1}{M-m} f(m) + \frac{1-m}{M-m} f(M) - \tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \tag{8.59}$$

$$\leq (1-m)(f[m,m] - f[m,M]) + \sum_{k=2}^{v-1} \frac{f^{(k)}(m)}{k!} \sum_{i=1}^{r} \frac{(p_i - mq_i)^k}{q_i^{k-1}}$$

$$+ \sum_{k=1}^{n-v} f[\underbrace{m,\ldots,m}_{v \text{ times}};\underbrace{M,\ldots,M}_{k \text{ times}}] \sum_{i=1}^{r} \frac{(p_i - mq_i)^v (p_i - mq_i)^{k-1}}{q_i^{v+k-2}}. \tag{8.60}$$

Inequality (8.60) also holds when the function f is n-concave and n and v are of equal parity. In case when the function f is n-convex and n and v are of equal parity, or when the function f is n-concave and n and v are of different parity, the inequality sign in (8.60) is reversed.

Proof. Let $\boldsymbol{x} = (x_1,\ldots,x_r)$ be such that $x_i \in [m,M]$ for $i = 1,\ldots,r$. In the relation (8.28) we can replace

$$g \longleftrightarrow \boldsymbol{x}, \text{ and } A(\boldsymbol{x}) = \sum_{i=1}^{r} p_i x_i.$$

In that way we get

$$\frac{M-\overline{x}}{M-m} f(m) + \frac{\overline{x}-m}{M-m} f(M) - \sum_{i=1}^{r} p_i f(x_i)$$

$$\leq (\overline{x}-m)(f[m,m] - f[m,M]) + \sum_{k=2}^{v-1} \frac{f^{(k)}(m)}{k!} \sum_{i=1}^{r} p_i (x_i - m)^k$$

$$+\sum_{k=1}^{n-v} f[\underbrace{m,\dots,m}_{v\ times};\underbrace{M,\dots,M}_{k\ times}]\sum_{i=1}^{r} p_i(x_i-m)^v(x_i-b)^{k-1},$$

where $\bar{x}=\sum_{i=1}^{n} p_i x_i$. In the previous relation we can set

$$p_i = q_i \text{ and } x_i = \frac{p_i}{q_i},$$

and after calculating

$$\bar{x}=\sum_{i=1}^{n} q_i\frac{p_i}{q_i}=\sum_{i=1}^{n} p_i = 1$$

we get (8.60). □

By utilizing Theorem 8.14 in the analogous way as above, we get an Edmundson-Lah-Ribarič type inequality for the generalized f-divergence functional (8.3) which does not depend on parity of n, and it is given in the following theorem.

Theorem 8.27 *Let $[m,M]\subset\mathbb{R}$ be an interval such that $m\le 1\le M$. Let $f\in\mathscr{C}^n([m,M])$ and let $\boldsymbol{p}=(p_1,\dots,p_r)$ and $\boldsymbol{p}=(q_1,\dots,q_r)$ be probability distributions such that $p_i/q_i\in[m,M]$ for every $i=1,\dots,r$. If the function f is n-convex and if $3\le v\le n-1$ is odd, then*

$$\begin{aligned}&\frac{M-1}{M-m}f(m)+\frac{1-m}{M-m}f(M)-\tilde{D}_f(\boldsymbol{p},\boldsymbol{q})\\ \le&(M-1)\left(f[m,M]-f[M,M]\right)+\sum_{k=2}^{v-1}\frac{f^{(k)}(M)}{k!}\sum_{i=1}^{r}\frac{(p_i-Mq_i)^k}{q_i^{k-1}}\\ &+\sum_{k=1}^{n-v} f[\underbrace{M,\dots,M}_{v\ times};\underbrace{m,\dots,m}_{k\ times}]\sum_{i=1}^{r}\frac{(p_i-Mq_i)^v(p_i-mq_i)^{k-1}}{q_i^{v+k-2}}\end{aligned}\tag{8.61}$$

Inequality (8.61) also holds when the function f is n-concave and v is even. In case when the function f is n-convex and v is even, or when the function f is n-concave and v is odd, the inequality sign in (8.61) is reversed.

Another generalization of the Edmundson-Lah-Ribarič inequality, which provides us with a lower and an upper bound for the generalized f-divergence functional, is given in the following theorem.

Theorem 8.28 *Let $[m,M]\subset\mathbb{R}$ be an interval such that $m\le 1\le M$. Let $f\in\mathscr{C}^n([m,M])$ and let $\boldsymbol{p}=(p_1,\dots,p_r)$ and $\boldsymbol{p}=(q_1,\dots,q_r)$ be probability distributions such that $p_i/q_i\in[m,M]$ for every $i=1,\dots,r$. If the function f is n-convex and if n is odd, then we have*

$$\begin{aligned}&\sum_{k=2}^{n-1} f[m;\underbrace{M,M,\dots,M}_{k\ times}]\sum_{i=1}^{r}\frac{(p_i-mq_i)(p_i-Mq_i)^{k-1}}{q_i^{k-1}}\le\frac{M-1}{M-m}f(m)+\frac{1-m}{M-m}f(M)-\tilde{D}_f(\boldsymbol{p},\boldsymbol{q})\\ \le& f[m,m;M]\sum_{i=1}^{r}\frac{(p_i-mq_i)(p_i-Mq_i)}{q_i}+\sum_{k=2}^{n-2} f[m,m;\underbrace{M,\dots,M}_{k\ times}]\sum_{i=1}^{r}\frac{(p_i-mq_i)^2(p_i-Mq_i)^{k-1}}{q_i^k}.\end{aligned}\tag{8.62}$$

Inequalities (8.62) also hold when the function f is n-concave and n is even. In case when the function f is n-convex and n is even, or when the function f is n-concave and n is odd, the inequality signs in (8.62) are reversed.

Proof. We start with inequalities (8.30) from Theorem 8.15, and follow the steps from the proof of Theorem 8.26. □

By utilizing Theorem 8.16 in an analogue way, we can get similar bounds for the generalized f-divergence functional that hold for all $n \in \mathbb{N}$, not only the odd ones.

Theorem 8.29 *Let $[m,M] \subset \mathbb{R}$ be an interval such that $m \leq 1 \leq M$. Let $f \in \mathscr{C}^n([m,M])$ and let $\boldsymbol{p} = (p_1,\ldots,p_r)$ and $\boldsymbol{p} = (q_1,\ldots,q_r)$ be probability distributions such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,r$. If the function f is n-convex, then we have*

$$f[M,M;m]\sum_{i=1}^{r}\frac{(p_i-mq_i)(p_i-Mq_i)}{q_i}+\sum_{k=2}^{n-2}f[M,M;\underbrace{m,m,\ldots,m}_{k\ times}]\sum_{i=1}^{r}\frac{(p_i-mq_i)^{k-1}(p_i-Mq_i)^2}{q_i^k}$$
$$\leq\frac{M-1}{M-m}f(m)+\frac{1-m}{M-m}f(M)-\tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \leq \sum_{k=2}^{n-1}f[M;\underbrace{m,\ldots,m}_{k\ times}]\sum_{i=1}^{r}\frac{(p_i-mq_i)^{k-1}(p_i-Mq_i)}{q_i^{k-1}}. \tag{8.63}$$

If the function f is n-concave, the inequality signs in (8.63) are reversed.

Remark 8.8 Let $\boldsymbol{p} = (p_1,\ldots,p_r)$ and $\boldsymbol{p} = (q_1,\ldots,q_r)$ be probability distributions.

- ▷ **Kullback-Leibler divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_{KL}(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{r} q_i \log\frac{q_i}{p_i},$$

and the corresponding generating function is $f(t) = t\log t, t > 0$. We can calculate

$$f^{(n)}(t) = (-1)^n(n-2)!t^{-(n-1)}.$$

It is clear that this function is $(2n-1)$-concave and $(2n)$-convex for any $n \in \mathbb{N}$.

- ▷ **Hellinger divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

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

and the corresponding generating function is $f(t) = \frac{1}{2}(1-\sqrt{t})^2, t > 0$. We see that

$$f^{(n)}(t) = (-1)^n\frac{(2n-3)!!}{2^n}t^{-\frac{2n-1}{2}},$$

so function f is $(2n-1)$-concave and $(2n)$-convex for any $n \in \mathbb{N}$.

▷ **Harmonic divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_{Ha}(\boldsymbol{p},\boldsymbol{q}) = \sum_{i=1}^{n} \frac{2p_iq_i}{p_i+q_i},$$

and the corresponding generating function is $f(t) = \frac{2t}{1+t}$. We can calculate

$$f^{(n)}(t) = 2(-1)^{n+1}n!(1+t)^{-(n+1)}.$$

Two cases need to be considered:

* if $t < -1$, then the function f is n-convex for every $n \in \mathbb{N}$;
* if $t > -1$, then the function f is $(2n)$-concave and $(2n-1)$-convex for any $n \in \mathbb{N}$.

▷ **Jeffreys divergence** of the probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$ is defined as

$$D_J(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{2}\sum_{i=1}^{n}(q_i-p_i)\log\frac{q_i}{p_i},$$

and the corresponding generating function is $f(t) = (1-t)\log\frac{1}{t}, t > 0$. After calculating, we see that

$$f^{(n)}(t) = (-1)^{n+1}t^{-n}(n-1)!(1+nt).$$

Obviously, this function is $(2n-1)$-convex and $(2n)$-concave for any $n \in \mathbb{N}$.

It is clear that all of the results from this section can be applied to the special types of divergences mentioned in this example.

8.4 Applications to Zipf-Mandelbrot law

Zipf-Mandelbrot law is a discrete probability distribution with parameters $N \in \mathbb{N}$, $q,s \in \mathbb{R}$ such that $q \geq 0$ and $s > 0$, possible values $\{1,2,\ldots,N\}$ and probability mass function

$$f(i;N,q,s) = \frac{1/(i+q)^s}{H_{N,q,s}}, \quad \text{where } H_{N,q,s} = \sum_{i=1}^{N}\frac{1}{(i+q)^s}.$$

It is used in various scientific fields: linguistics [29], information sciences [12, 34], ecological field studies [30] and music [23]. Benoit Mandelbrot in 1966 gave improvement of Zipf law for the count of the low-rank words. Various scientific fields use this law for different purposes, for example information sciences use it for indexing [12, 34], ecological field studies in predictability of ecosystem [30], in music is used to determine aesthetically pleasing music [23].

Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively. We can use Corollary 8.2 and Corollary 8.3 in a similar way as described above in order to obtain inequalities for the Kullback-Leibler divergence. Let us denote

$$\begin{aligned} H_{N,q_1,s_1} &= H_1,\ H_{N,q_2,s_2} = H_2 \\ m_{\boldsymbol{p},\boldsymbol{q}} &:= \min\left\{\frac{p_i}{q_i}\right\} = \frac{H_2}{H_1}\min\left\{\frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}}\right\} \\ M_{\boldsymbol{p},\boldsymbol{q}} &:= \max\left\{\frac{p_i}{q_i}\right\} = \frac{H_2}{H_1}\max\left\{\frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}}\right\} \end{aligned} \tag{8.64}$$

Corollary 8.8 ([25]) *Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively. If the base of the logarithm is greater than one, we have*

$$\begin{aligned} 0 &\leq D_{KL}(\boldsymbol{p},\boldsymbol{q}) \\ &\leq (M_{\boldsymbol{p},\boldsymbol{q}}-1)(1-m_{\boldsymbol{p},\boldsymbol{q}}) \sup_{t\in\langle m_{\boldsymbol{p},\boldsymbol{q}},M_{\boldsymbol{p},\boldsymbol{q}}\rangle} \Psi_{id\cdot\log}(t;m_{\boldsymbol{p},\boldsymbol{q}},M_{P,Q}) - \Delta_{\boldsymbol{p},\boldsymbol{q}} \\ &\leq \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}(M_{\boldsymbol{p},\boldsymbol{q}}-1)(1-m_{\boldsymbol{p},\boldsymbol{q}})\log\frac{M_{\boldsymbol{p},\boldsymbol{q}}}{m_{\boldsymbol{p},\boldsymbol{q}}} - \Delta_{\boldsymbol{p},\boldsymbol{q}} \\ &\leq \frac{1}{4}(M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}})\log\frac{M_{\boldsymbol{p},\boldsymbol{q}}}{m_{\boldsymbol{p},\boldsymbol{q}}} - \Delta_{\boldsymbol{p},\boldsymbol{q}} \end{aligned} \tag{8.65}$$

and

$$\begin{aligned} \Delta_{\boldsymbol{p},\boldsymbol{q}} \leq& \frac{M_{\boldsymbol{p},\boldsymbol{q}}m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}\log\left(\frac{m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}}}\right) + \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}\log\left(\frac{M_{\boldsymbol{p},\boldsymbol{q}}^{M_{\boldsymbol{p},\boldsymbol{q}}}}{m_{\boldsymbol{p},\boldsymbol{q}}^{m_{\boldsymbol{p},\boldsymbol{q}}}}\right) - D_{KL}(\boldsymbol{p},\boldsymbol{q}) \\ \leq& (M_{\boldsymbol{p},\boldsymbol{q}}-1)(1-m_{\boldsymbol{p},\boldsymbol{q}}) \sup_{t\in\langle m_{\boldsymbol{p},\boldsymbol{q}},M_{\boldsymbol{p},\boldsymbol{q}}\rangle} \Psi_{id\cdot\log}(t;m_{\boldsymbol{p},\boldsymbol{q}},M_{\boldsymbol{p},\boldsymbol{q}}) \\ \leq& \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}(M_{\boldsymbol{p},\boldsymbol{q}}-1)(1-m_{\boldsymbol{p},\boldsymbol{q}})\log\left(\frac{M_{\boldsymbol{p},\boldsymbol{q}}}{m_{\boldsymbol{p},\boldsymbol{q}}}\right) \leq \frac{1}{4}(M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}})\log\left(\frac{M_{\boldsymbol{p},\boldsymbol{q}}}{m_{\boldsymbol{p},\boldsymbol{q}}}\right), \end{aligned} \tag{8.66}$$

where $D_{KL}(\boldsymbol{p},\boldsymbol{q})$ is the Kullback-Leibler divergence of distributions $\boldsymbol{p}$ and $\boldsymbol{q}$, $m_{\boldsymbol{p},\boldsymbol{q}}$ and $M_{\boldsymbol{p},\boldsymbol{q}}$ are defined in (8.64), and

$$\begin{aligned} \Delta_{\boldsymbol{p},\boldsymbol{q}} =& \left(\frac{1}{2} - \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}\sum_{i=1}^{N}\left|\frac{1}{H_1(i+q_1)^{s_1}} - \frac{m_{\boldsymbol{p},\boldsymbol{q}}+M_{\boldsymbol{p},\boldsymbol{q}}}{2}\cdot\frac{1}{H_2(i+q_2)^{s_2}}\right|\right) \\ &\times\left(m_{\boldsymbol{p},\boldsymbol{q}}\log\frac{2m_{\boldsymbol{p},\boldsymbol{q}}}{m_{\boldsymbol{p},\boldsymbol{q}}+M_{\boldsymbol{p},\boldsymbol{q}}} + M_{\boldsymbol{p},\boldsymbol{q}}\log\frac{2M_{\boldsymbol{p},\boldsymbol{q}}}{m_{\boldsymbol{p},\boldsymbol{q}}+M_{\boldsymbol{p},\boldsymbol{q}}}\right) \end{aligned}$$

Remark 8.9 If we utilize Remark 8.4 and Remark 8.5 in the same way as described above, we can obtain companion inequalities for the reversed Kullback-Leibler divergence $D_{KL}(\boldsymbol{q},\boldsymbol{p})$ of these distributions.

For finite N and $q = 0$ the Zipf-Mandelbrot law becomes Zipf's law. I is one of the basic laws in information science and bibliometrics, but it is also often used in linguistics. George Zipf's in 1932 found that we can count how many times each word appears in the text. So if we ranked (r) word according to the frequency of word occurrence (f), the product of these two numbers is a constant $C = r * f$. Same law in mathematical sense is also used in other scientific disciplines, but name of the law can be different, since regularities in different scientific fields are discovered independently from each other. In economics same law or regularity are called Pareto's law which analyze and predicts the distribution of the wealthiest members of the community [10]. The same type of distribution that we have in Zipf's and Pareto's law, also known as the Power law, can be found in wide variety of scientific disciplines, such as: physics, biology, earth and planetary sciences, computer science, demography and the social sciences [31] and many others. At this point of time we will not explain usage and their importance of this law in each scientific field, but we will retain on frequency of the word usage. Since, words are one of basic properties in human communication system. That frequency of used word and human communication system can be explained with plain mathematical formula is extremely interesting and useful in analysis of language and their usage. Since this law is be applicable in indexing and text mining, it is quite useful in today's world in which we use Internet to retrive most of the information that we need.

Probability mass function of Zipf's law is:

$$f(k;N,s) = \frac{1/k^s}{H_{N,s}}, \quad \text{where} \quad H_{N,s} = \sum_{i=1}^{N} \frac{1}{i^s}.$$

Since Zipf's law is a special case of the Zipf-Mandelbrot law, all of the results from above hold for $q = 0$.

Gelbukh and Sidorov in [13] observed the difference between the coefficients s_1 and s_2 in Zipf's law for the russian and english language. They processed 39 literature texts for each language, chosen randomly from different genres, with the requirement that the size be greater than 10,000 running words each. They calculated coefficients for each of the mentioned texts and as the result they obtained the average of $s_1 = 0,892869$ for the russian language, and $s_2 = 0,973863$ for the english language.

If we take $q_1 = q_2 = 0$, we can use the results from the above regarding the Kullback-Leibler divergence of two Zipf-Mandelbrot distributions in order to give estimates for the Kullback-Leibler divergence of the distributions associated to the russian and english language. For those experimental values of s_1 and s_2 we have

$$m_N = \min\left\{\frac{p_i}{q_i}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}} \min\left\{\frac{i^{s_2}}{i^{s_1}}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}} \min\left\{i^{s_2-s_1}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}}$$

and

$$M_N = \max\left\{\frac{p_i}{q_i}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}} \max\left\{\frac{i^{s_2}}{i^{s_1}}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}} \max\left\{i^{s_2-s_1}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}} N^{0,080994}.$$

Hence the following bounds for the mentioned divergence, arising from Corollary 8.8 and depending only on the parameter N, hold.

$$\begin{aligned}
0 &\leq D_{KL}(\boldsymbol{p},\boldsymbol{q}) \\
&\leq (M_N-1)(1-m_N) \sup_{t\in\langle m_N,M_N\rangle} \Psi_{id\cdot\log}(t;m_N,M_N) - \Delta_N \\
&\leq \frac{0,080994}{M_N-m_N}(M_N-1)(1-m_N)\log N - \Delta_N \\
&\leq 0,020249(M_N-m_N)\log N - \Delta_N
\end{aligned}$$

We also have

$$\begin{aligned}
\Delta_N \leq& \frac{0,080994N^{0,080994}}{N^{0,080994}-1}\left(1-\frac{H_{N;0,973863}}{H_{N;0,892869}}\right)\log N + \log\left(\frac{H_{N;0,973863}}{H_{N;0,892869}}\right) - D_{KL}(\boldsymbol{p},\boldsymbol{q}) \\
\leq& (M_N-1)(1-m_N) \sup_{t\in\langle m_N,M_N\rangle} \Psi_{id\cdot\log}(t;m_N,M_N) \\
\leq& \frac{0,080994}{M_N-m_N}(M_N-1)(1-m_N)\log N \leq 0,020249(M_N-m_N)\log N,
\end{aligned}$$

where

$$\begin{aligned}
\Delta_N =& \left(\frac{1}{2} - \frac{1}{H_{N;0,973863}(N^{0,080994}-1)}\sum_{i=1}^{N}\left|\frac{1}{i^{0,892869}} - \frac{N^{0,080994}+1}{2i^{0,973863}}\right|\right) \\
&\times\left(\log\frac{2}{N^{0,080994}+1} + N^{0,080994}\log\frac{2N^{0,080994}}{N^{0,080994}+1}\right)\frac{H_{N;0,973863}}{H_{N;0,892869}}
\end{aligned}$$

By calculating the above results for the Kullback-Leibler divergence of the distributions associated to the russian ($\boldsymbol{p}$) and english ($\boldsymbol{q}$) language for different values of the parameter N, we obtained the following bounds:

- from the first series of inequalities:

N	5000	10000	50000	100000
$D_{KL}(\boldsymbol{p},\boldsymbol{q}) \leq$	0,0862934	0,100855	0,138862	0,157016

- from the second series of inequalities:

N	5000	10000	50000	100000
$D_{KL}(\boldsymbol{p},\boldsymbol{q}) \leq$	0,00106	0,001274	0,0018269	0,002091

The base of the logarithm used in our calculations is 2.

Again, $\boldsymbol{p}$ and $\boldsymbol{q}$ are Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively, and let $m_{\boldsymbol{p},\boldsymbol{q}}$ and $M_{\boldsymbol{p},\boldsymbol{q}}$ be defined in (8.64).

Next result is a special case of Corollary 8.4 and Corollary 8.5, and it gives us Edmundson-Lah-Ribarič and Jensen type inequalities for the Kullback-Leibler divergence of two Zipf-Mandelbrot laws. In contrast to previous results, function f is not necessarily convex in the classical sense.

Corollary 8.9 ([26]) Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively. Then we have

$$\begin{aligned}
0 \leq & \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \left(\log M_{\boldsymbol{p},\boldsymbol{q}} + 1 - \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \log \frac{M_{\boldsymbol{p},\boldsymbol{q}}^{M_{\boldsymbol{p},\boldsymbol{q}}}}{m_{\boldsymbol{p},\boldsymbol{q}}^{m_{\boldsymbol{p},\boldsymbol{q}}}} \right) \times \\
& \sum_{i=1}^{N} \frac{1}{(i+q_2)^{s_2} H_2} \left(M_{\boldsymbol{p},\boldsymbol{q}} - \frac{H_2}{H_1} \frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}} \right) \left(\frac{H_2}{H_1} \frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}} \right) \\
\leq & \frac{m_{\boldsymbol{p},\boldsymbol{q}} M_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \log \frac{m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}}} + \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \log \frac{M_{\boldsymbol{p},\boldsymbol{q}}^{M_{\boldsymbol{p},\boldsymbol{q}}}}{m_{\boldsymbol{p},\boldsymbol{q}}^{m_{\boldsymbol{p},\boldsymbol{q}}}} - D_{KL}(\boldsymbol{p},\boldsymbol{q}) \\
\leq & \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \left(\frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \log \frac{M_{\boldsymbol{p},\boldsymbol{q}}^{M_{\boldsymbol{p},\boldsymbol{q}}}}{m_{\boldsymbol{p},\boldsymbol{q}}^{m_{\boldsymbol{p},\boldsymbol{q}}}} - \log m_{\boldsymbol{p},\boldsymbol{q}} - 1 \right) \times \\
& \sum_{i=1}^{N} \frac{1}{(i+q_2)^{s_2} H_2} \left(M_{\boldsymbol{p},\boldsymbol{q}} - \frac{H_2}{H_1} \frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}} \right) \left(\frac{H_2}{H_1} \frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}} \right) \\
\leq & \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \left(\frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \log \frac{M_{\boldsymbol{p},\boldsymbol{q}}^{M_{\boldsymbol{p},\boldsymbol{q}}}}{m_{\boldsymbol{p},\boldsymbol{q}}^{m_{\boldsymbol{p},\boldsymbol{q}}}} - \log m_{\boldsymbol{p},\boldsymbol{q}} - 1 \right) (M_{\boldsymbol{p},\boldsymbol{q}} - 1)(1 - m_{\boldsymbol{p},\boldsymbol{q}}) \\
\leq & \frac{1}{4} \left(\log \frac{M_{\boldsymbol{p},\boldsymbol{q}}^{M_{\boldsymbol{p},\boldsymbol{q}}}}{m_{\boldsymbol{p},\boldsymbol{q}}^{m_{\boldsymbol{p},\boldsymbol{q}}}} - (\log m_{\boldsymbol{p},\boldsymbol{q}} + 1)(M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}) \right)
\end{aligned}$$

and

$$\begin{aligned}
0 \leq & D_{KL}(\boldsymbol{p},\boldsymbol{q}) \leq \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \left[\left(\frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \log \frac{M_{\boldsymbol{p},\boldsymbol{q}}^{M_{\boldsymbol{p},\boldsymbol{q}}}}{m_{\boldsymbol{p},\boldsymbol{q}}^{m_{\boldsymbol{p},\boldsymbol{q}}}} - \log m - 1 \right) (M_{\boldsymbol{p},\boldsymbol{q}} - 1)(1 - m_{\boldsymbol{p},\boldsymbol{q}}) \right. \\
& - \left(\log M_{\boldsymbol{p},\boldsymbol{q}} + 1 - \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} \log \frac{M_{\boldsymbol{p},\boldsymbol{q}}^{M_{\boldsymbol{p},\boldsymbol{q}}}}{m_{\boldsymbol{p},\boldsymbol{q}}^{m_{\boldsymbol{p},\boldsymbol{q}}}} \right) \times \\
& \left. \sum_{i=1}^{N} \frac{1}{(i+q_2)^{s_2} H_2} \left(M_{\boldsymbol{p},\boldsymbol{q}} - \frac{H_2}{H_1} \frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}} \right) \left(\frac{H_2}{H_1} \frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}} \right) \right].
\end{aligned}$$

Remark 8.10 From Corollary 8.6 and Corollary 8.7 we can obtain the same type of inequalities, but for the reversed Kullback-Leibler divergence $D_{KL}(\boldsymbol{q},\boldsymbol{p})$ of the Zipf-Mandelbrot distributions $\boldsymbol{p}$ and $\boldsymbol{q}$.

Since Zipf's law is a special case of the Zipf-Mandelbrot law, two previous results hold for Zipf's law with $q = 0$.

Again, if we take $q_1 = q_2 = 0$, we can use the results from the above regarding the Kullback-Leibler divergence of two Zipf-Mandelbrot distributions in order to give estimates for the Kullback-Leibler divergence of the distributions associated to the russian and english language. As said before, for those experimental values of s_1 and s_2 we have

$$m_N = \min \left\{ \frac{p_i}{q_i} \right\} = \frac{H_{N,s_2}}{H_{N,s_1}} \min \left\{ \frac{i^{s_2}}{i^{s_1}} \right\} = \frac{H_{N,s_2}}{H_{N,s_1}} \min \left\{ i^{s_2 - s_1} \right\} = \frac{H_{N,s_2}}{H_{N,s_1}}$$

and

$$M_N = \max\left\{\frac{p_i}{q_i}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}}\max\left\{\frac{i^{s_2}}{i^{s_1}}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}}\max\left\{i^{s_2-s_1}\right\} = \frac{H_{N,s_2}}{H_{N,s_1}}N^{0,080994}.$$

Hence the following bounds for the mentioned divergence, depending only on the parameter N, hold.

$$\begin{aligned}
0 \leq & \frac{1}{M_N-m_N}\left(\log M_N + 1 - \frac{1}{M_N-m_N}\log\frac{M_N^{M_N}}{m_N^{m_N}}\right)\times \\
& \sum_{i=1}^{N}\frac{H_{N,0,973863}}{i^{0,973863}H_{N,0,892869}^2}\left(N^{0,080994}-i^{0,080994}\right)\left(i^{0,080994}-1\right) \\
\leq & \frac{m_N M_N}{M_N-m_N}\log\frac{m_N}{M_N} + \frac{1}{M_N-m_N}\log\frac{M_N^{M_N}}{m_N^{m_N}} - D_{KL}(\boldsymbol{p},\boldsymbol{q}) \\
\leq & \frac{1}{M_N-m_N}\left(\frac{1}{M_N-m_N}\log\frac{M_N^{M_N}}{m_N^{m_N}} - \log m_N - 1\right)\times \\
& \sum_{i=1}^{N}\frac{H_{N,0,973863}}{i^{0,973863}H_{N,0,892869}^2}\left(N^{0,080994}-i^{0,080994}\right)\left(i^{0,080994}-1\right) \\
\leq & \frac{1}{M_N-m_N}\left(\frac{1}{M_N-m_N}\log\frac{M_N^{M_N}}{m_N^{m_N}} - \log m_N - 1\right)(M_N-1)(1-m_N) \\
\leq & \frac{1}{4}\left(\log\frac{M_N^{M_N}}{m_N^{m_N}} - (\log m_N + 1)(M_N-m_N)\right)
\end{aligned}$$

$$\begin{aligned}
0 \leq D_{KL}(\boldsymbol{p},\boldsymbol{q}) \leq & \frac{1}{M_N-m_N}\left[\left(\frac{1}{M_N-m_N}\log\frac{M_N^{M_N}}{m_N^{m_N}} - \log m_N - 1\right)(M_N-1)(1-m_N)\right. \\
& -\left(\log M + 1 - \frac{1}{M-m}\log\frac{M^M}{m^m}\right)\times \\
& \left.\sum_{i=1}^{N}\frac{H_{N,0,973863}}{i^{0,973863}H_{N,0,892869}^2}\left(N^{0,080994}-i^{0,080994}\right)\left(i^{0,080994}-1\right)\right].
\end{aligned}$$

By calculating the above results for the Kullback-Leibler divergence of the distributions associated to the russian ($\boldsymbol{p}$) and english ($\boldsymbol{q}$) language for different values of the parameter N, we obtained the following bounds:

- from the first series of inequalities:

N	5000	10000	50000	100000
$D_{KL}(\boldsymbol{p},\boldsymbol{q}) \leq$	1.19101	1.16826	1.12176	1.10408

- from the second series of inequalities:

N	5000	10000	50000	100000
$D_{KL}(\boldsymbol{p},\boldsymbol{q}) \leq$	0.170194	0.189118	0.236439	0.258335

The base of the logarithm used in our calculations is 2.

The result that follows is a special case of Theorem 8.22, and it gives us Edmundson-Lah-Ribarič type inequality for the generalized f-divergence of the Zipf-Mandelbrot law.

Corollary 8.10 ([24]) *Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively, and let $m_{\boldsymbol{p},\boldsymbol{q}}$ and $M_{\boldsymbol{p},\boldsymbol{q}}$ be defined in (8.64). Let $f\colon [m_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}] \to \mathbb{R}$ be a 3-convex function. Then we have*

$$\begin{aligned}
&\frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\left(\frac{f(M_{\boldsymbol{p},\boldsymbol{q}}) - f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} - f'_{+}(m_{\boldsymbol{p},\boldsymbol{q}})\right) \times \\
&\quad \sum_{i=1}^{n} \frac{1}{(i+q_2)^{s_2} H_2}\left(M_{\boldsymbol{p},\boldsymbol{q}} - \frac{(i+q_2)^{s_2} H_2}{(i+q_1)^{s_1} H_1}\right)\left(\frac{(i+q_2)^{s_2} H_2}{(i+q_1)^{s_1} H_1} - m_{\boldsymbol{p},\boldsymbol{q}}\right) \\
\leq &\frac{M_{\boldsymbol{p},\boldsymbol{q}} - 1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} f(m_{\boldsymbol{p},\boldsymbol{q}}) + \frac{1 - m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} f(M_{\boldsymbol{p},\boldsymbol{q}}) - \tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \\
\leq &\frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\left(f'_{-}(M_{\boldsymbol{p},\boldsymbol{q}}) - \frac{f(M_{\boldsymbol{p},\boldsymbol{q}}) - f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\right) \times \\
&\quad \sum_{i=1}^{n} \frac{1}{(i+q_2)^{s_2} H_2}\left(M_{\boldsymbol{p},\boldsymbol{q}} - \frac{(i+q_2)^{s_2} H_2}{(i+q_1)^{s_1} H_1}\right)\left(\frac{(i+q_2)^{s_2} H_2}{(i+q_1)^{s_1} H_1} - m_{\boldsymbol{p},\boldsymbol{q}}\right).
\end{aligned} \tag{8.67}$$

Next result follows directly from Theorem 8.23, and it represents a Jensen type inequality for the generalized f-divergence of the Zipf-Mandelbrot law without the assumption on the convexity of the function f in the classical sense.

Corollary 8.11 ([24]) *Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively, and let $m_{\boldsymbol{p},\boldsymbol{q}}$ and $M_{\boldsymbol{p},\boldsymbol{q}}$ be defined in (8.64). Let $f\colon [m_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}] \to \mathbb{R}$ be a 3-convex function. Then we have*

$$\begin{aligned}
&\frac{(M_{\boldsymbol{p},\boldsymbol{q}} - 1)(1 - m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\left(\frac{f(M_{\boldsymbol{p},\boldsymbol{q}}) - f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} - f'_{+}(m_{\boldsymbol{p},\boldsymbol{q}})\right) \\
&\quad - \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\left(f'_{-}(M_{\boldsymbol{p},\boldsymbol{q}}) - \frac{f(M_{\boldsymbol{p},\boldsymbol{q}}) - f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\right) \times \\
&\quad \sum_{i=1}^{n} \frac{1}{(i+q_2)^{s_2} H_2}\left(M_{\boldsymbol{p},\boldsymbol{q}} - \frac{(i+q_2)^{s_2} H_2}{(i+q_1)^{s_1} H_1}\right)\left(\frac{(i+q_2)^{s_2} H_2}{(i+q_1)^{s_1} H_1} - m_{\boldsymbol{p},\boldsymbol{q}}\right) \\
\leq &\tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) - f(1) \\
\leq &\frac{(M_{\boldsymbol{p},\boldsymbol{q}} - 1)(1 - m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\left(f'_{-}(M_{\boldsymbol{p},\boldsymbol{q}}) - \frac{f(M_{\boldsymbol{p},\boldsymbol{q}}) - f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\right) \\
&\quad - \frac{1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}}\left(\frac{f(M_{\boldsymbol{p},\boldsymbol{q}}) - f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} - f'_{+}(m_{\boldsymbol{p},\boldsymbol{q}})\right) \times
\end{aligned} \tag{8.68}$$

$$\sum_{i=1}^{n}\frac{1}{(i+q_2)^{s_2}H_2}\left(M_{\boldsymbol{p},\boldsymbol{q}}-\frac{(i+q_2)^{s_2}H_2}{(i+q_1)^{s_1}H_1}\right)\left(\frac{(i+q_2)^{s_2}H_2}{(i+q_1)^{s_1}H_1}-m_{\boldsymbol{p},\boldsymbol{q}}\right).$$

Remark 8.11 Corollary 8.10 and Corollary 8.11 can easily be applied to Kullback-Leibler divergence, Hellinger divergence, Renyi divergence, harmonic divergence or Jeffreys divergence considering Remark 8.7.

The following result is a special case of Theorem 8.24, and it gives us Edmundson-Lah-Ribarič type inequality for the generalized f-divergence of the Zipf-Mandelbrot law.

Corollary 8.12 ([27]) *Let* $\boldsymbol{p}$ *and* $\boldsymbol{q}$ *be Zipf-Mandelbrot laws with parameters* $N\in\mathbb{N}$, $q_1,q_2\geq 0$ *and* $s_1,s_2>0$ *respectively, and let* $m_{\boldsymbol{p},\boldsymbol{q}}$ *and* $M_{\boldsymbol{p},\boldsymbol{q}}$ *be defined in (8.64). Let* $f\colon[m_{\boldsymbol{p},\boldsymbol{q}},M_{\boldsymbol{p},\boldsymbol{q}}]\to\mathbb{R}$ *be a 3-convex function. Then we have*

$$\begin{aligned}
&(1-m_{\boldsymbol{p},\boldsymbol{q}})\left[\frac{f(M_{\boldsymbol{p},\boldsymbol{q}})-f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}-\frac{f'_+(m_{\boldsymbol{p},\boldsymbol{q}})}{2}\right]\\
&\quad-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{1}{(i+q_1)^{s_1}H_1}-\frac{m_{\boldsymbol{p},\boldsymbol{q}}}{(i+q_2)^{s_2}H_2}\right)f'\left(\frac{H_2}{H_1}\frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}}\right)\\
\leq&\frac{M_{\boldsymbol{p},\boldsymbol{q}}-1}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}f(m_{\boldsymbol{p},\boldsymbol{q}})+\frac{1-m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}f(M_{\boldsymbol{p},\boldsymbol{q}})-\tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \qquad (8.69)\\
\leq&\frac{1}{2}\sum_{i=1}^{n}\left(\frac{M_{\boldsymbol{p},\boldsymbol{q}}}{(i+q_2)^{s_2}H_2}-\frac{1}{(i+q_1)^{s_1}H_1}\right)f'\left(\frac{H_2}{H_1}\frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}}\right)\\
&\quad-(M_{\boldsymbol{p},\boldsymbol{q}}-1)\left[\frac{f(M_{\boldsymbol{p},\boldsymbol{q}})-f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}-\frac{f'_-(M_{\boldsymbol{p},\boldsymbol{q}})}{2}\right].
\end{aligned}$$

Next result follows directly from Theorem 8.12, and it gives us another Edmundson-Lah-Ribarič type inequality for the generalized f-divergence of the Zipf-Mandelbrot law.

Corollary 8.13 ([27]) *Let* $\boldsymbol{p}$ *and* $\boldsymbol{q}$ *be Zipf-Mandelbrot laws with parameters* $N\in\mathbb{N}$, $q_1,q_2\geq 0$ *and* $s_1,s_2>0$ *respectively, and let* $m_{\boldsymbol{p},\boldsymbol{q}}$ *and* $M_{\boldsymbol{p},\boldsymbol{q}}$ *be defined in (8.64). Let* $f\colon[m_{\boldsymbol{p},\boldsymbol{q}},M_{\boldsymbol{p},\boldsymbol{q}}]\to\mathbb{R}$ *be a 3-convex function. Then we have*

$$\begin{aligned}
&(M_{\boldsymbol{p},\boldsymbol{q}}-1)\left[f'_-(M_{\boldsymbol{p},\boldsymbol{q}})-\frac{f(M_{\boldsymbol{p},\boldsymbol{q}})-f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}\right]\\
&\quad-\frac{f''_-(M_{\boldsymbol{p},\boldsymbol{q}})}{2}\sum_{i=1}^{n}(i+q_2)^{s_2}H_2\left(\frac{M_{\boldsymbol{p},\boldsymbol{q}}}{(i+q_2)^{s_2}H_2}-\frac{1}{(i+q_1)^{s_1}H_1}\right)^2\\
\leq&\frac{M_{\boldsymbol{p},\boldsymbol{q}}-1}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}f(m_{\boldsymbol{p},\boldsymbol{q}})+\frac{1-m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}f(M_{\boldsymbol{p},\boldsymbol{q}})-\tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \qquad (8.70)\\
\leq&(1-m_{\boldsymbol{p},\boldsymbol{q}})\left[\frac{f(M_{\boldsymbol{p},\boldsymbol{q}})-f(m_{\boldsymbol{p},\boldsymbol{q}})}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}}-f'_+(m_{\boldsymbol{p},\boldsymbol{q}})\right]\\
&\quad-\frac{f''_+(m_{\boldsymbol{p},\boldsymbol{q}})}{2}\sum_{i=1}^{n}(i+q_2)^{s_2}H_{N,q_2,s_2}\left(\frac{1}{(i+q_1)^{s_1}H_1}-\frac{m_{\boldsymbol{p},\boldsymbol{q}}}{(i+q_2)^{s_2}H_2}\right)^2.
\end{aligned}$$

Remark 8.12 Again, by taking into consideration Remark 8.7 one can see that Corollary 8.12 and Corollary 8.13 can easily be applied to any of the following divergences: Kullback-Leibler divergence, Hellinger divergence, Renyi divergence, harmonic divergence or Jeffreys divergence.

The following results are special cases of Theorems 8.26, 8.27, 8.28 and 8.29 respectively, and they gives us Edmundson-Lah-Ribarič type inequality for the generalized f-divergence of the Zipf-Mandelbrot law.

Corollary 8.14 ([28]) Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively, and let H_1, H_2, $m_{\boldsymbol{p},\boldsymbol{q}}$ and $M_{\boldsymbol{p},\boldsymbol{q}}$ be defined in (8.64). Let Let $f \in \mathscr{C}^n([m_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}])$ be a n-convex function. If n and $3 \leq v \leq n-1$ are of different parity, then

$$
\begin{aligned}
&\frac{M_{\boldsymbol{p},\boldsymbol{q}}-1}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}} f(m_{\boldsymbol{p},\boldsymbol{q}}) + \frac{1-m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}} f(M_{\boldsymbol{p},\boldsymbol{q}}) - \tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \\
&\leq (1-m_{\boldsymbol{p},\boldsymbol{q}}) \left(f'(m_{\boldsymbol{p},\boldsymbol{q}}) - f[m_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}]\right) + \sum_{k=2}^{v-1} \frac{f^{(k)}(m_{\boldsymbol{p},\boldsymbol{q}})}{H_2 k!} \sum_{i=1}^{r} \frac{\left(\frac{H_2 (i+q_2)^{s_2}}{H_1 (i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)^k}{(i+q_2)^{s_2}} \\
&\quad + \sum_{k=1}^{n-v} f[\underbrace{m_{\boldsymbol{p},\boldsymbol{q}}, \ldots, m_{\boldsymbol{p},\boldsymbol{q}}}_{v \text{ times}}; \underbrace{M_{\boldsymbol{p},\boldsymbol{q}}, \ldots, M_{\boldsymbol{p},\boldsymbol{q}}}_{k \text{ times}}] \sum_{i=1}^{r} \frac{\left(\frac{H_2 (i+q_2)^{s_2}}{H_1 (i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)^m \left(\frac{H_2 (i+q_2)^{s_2}}{H_1 (i+q_1)^{s_1}} - M_{\boldsymbol{p},\boldsymbol{q}}\right)^{k-1}}{H_2 (i+q_2)^{s_2}}.
\end{aligned}
$$

This inequality also holds when the function f is n-concave and n and v are of equal parity. In case when the function f is n-convex and n and v are of equal parity, or when the function f is n-concave and n and v are of different parity, the inequality sign is reversed.

Corollary 8.15 ([28]) Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively, and let H_1, H_2, $m_{\boldsymbol{p},\boldsymbol{q}}$ and $M_{\boldsymbol{p},\boldsymbol{q}}$ be defined in (8.64). Let Let $f \in \mathscr{C}^n([m_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}])$ be a n-convex function. If the function f is n-convex and if $3 \leq v \leq n-1$ are of different parity, then

$$
\begin{aligned}
&\frac{M_{\boldsymbol{p},\boldsymbol{q}}-1}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}} f(m_{\boldsymbol{p},\boldsymbol{q}}) + \frac{1-m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}}-m_{\boldsymbol{p},\boldsymbol{q}}} f(M_{\boldsymbol{p},\boldsymbol{q}}) - \tilde{D}_f(\boldsymbol{p},\boldsymbol{q}) \\
&\leq (M_{\boldsymbol{p},\boldsymbol{q}}-1) \left(f[m_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}] - f'(M_{\boldsymbol{p},\boldsymbol{q}})\right) + \sum_{k=2}^{v-1} \frac{f^{(k)}(M_{\boldsymbol{p},\boldsymbol{q}})}{H_2 k!} \sum_{i=1}^{r} \frac{\left(\frac{H_2 (i+q_2)^{s_2}}{H_1 (i+q_1)^{s_1}} - b_{\boldsymbol{p},\boldsymbol{q}}\right)^k}{(i+q_2)^{s_2}} \\
&\quad + \sum_{k=1}^{n-v} f[\underbrace{M_{\boldsymbol{p},\boldsymbol{q}}, \ldots, M_{\boldsymbol{p},\boldsymbol{q}}}_{v \text{ times}}; \underbrace{m_{\boldsymbol{p},\boldsymbol{q}}, \ldots, m_{\boldsymbol{p},\boldsymbol{q}}}_{k \text{ times}}] \sum_{i=1}^{r} \frac{\left(\frac{H_2 (i+q_2)^{s_2}}{H_1 (i+q_1)^{s_1}} - M_{\boldsymbol{p},\boldsymbol{q}}\right)^v \left(\frac{H_2 (i+q_2)^{s_2}}{H_1 (i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)^{k-1}}{H_2 (i+q_2)^{s_2}}.
\end{aligned}
$$

The inequality above also holds when the function f is n-concave and v is even. In case when the function f is n-convex and v is even, or when the function f is n-concave and v is odd, the inequality sign is reversed.

Corollary 8.16 ([28]) Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively, and let H_1, H_2, $m_{\boldsymbol{p},\boldsymbol{q}}$ and $M_{\boldsymbol{p},\boldsymbol{q}}$ be defined in (8.64). Let Let $f \in \mathscr{C}^n([m_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}])$ be a n-convex function. If the function f is n-convex and if n is odd, then we have

$$\sum_{k=2}^{n-1} f[m_{\boldsymbol{p},\boldsymbol{q}}; \underbrace{M_{\boldsymbol{p},\boldsymbol{q}}, \ldots, M_{\boldsymbol{p},\boldsymbol{q}}}_{k \text{ times}}] \sum_{i=1}^{r} \frac{\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - M_{\boldsymbol{p},\boldsymbol{q}}\right)^{k-1}}{H_2(i+q_2)^{s_2}}$$
$$\leq \frac{M_{\boldsymbol{p},\boldsymbol{q}} - 1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} f(m_{\boldsymbol{p},\boldsymbol{q}}) + \frac{1 - m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} f(M_{\boldsymbol{p},\boldsymbol{q}}) - \tilde{D}_f(\boldsymbol{p},\boldsymbol{q})$$
$$\leq f[m_{\boldsymbol{p},\boldsymbol{q}}, m_{\boldsymbol{p},\boldsymbol{q}}; M_{\boldsymbol{p},\boldsymbol{q}}] \sum_{i=1}^{r} \frac{\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - M_{\boldsymbol{p},\boldsymbol{q}}\right)}{H_2(i+q_2)^{s_2}}$$
$$+ \sum_{k=2}^{n-2} f[m_{\boldsymbol{p},\boldsymbol{q}}, m_{\boldsymbol{p},\boldsymbol{q}}; \underbrace{M_{\boldsymbol{p},\boldsymbol{q}}, \ldots, M_{\boldsymbol{p},\boldsymbol{q}}}_{k \text{ times}}] \sum_{i=1}^{r} \frac{\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)^2 \left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - M_{\boldsymbol{p},\boldsymbol{q}}\right)^{k-1}}{H_2(i+q_2)^{s_2}}.$$

These inequalities also hold when the function f is n-concave and n is even. In case when the function f is n-convex and n is even, or when the function f is n-concave and n is odd, the inequality signs are reversed.

Corollary 8.17 ([28]) Let $\boldsymbol{p}$ and $\boldsymbol{q}$ be Zipf-Mandelbrot laws with parameters $N \in \mathbb{N}$, $q_1, q_2 \geq 0$ and $s_1, s_2 > 0$ respectively, and let H_1, H_2, $m_{\boldsymbol{p},\boldsymbol{q}}$ and $M_{\boldsymbol{p},\boldsymbol{q}}$ be defined in (8.64). Let Let $f \in \mathscr{C}^n([m_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}])$ be a n-convex function. If the function f is n-convex, then we have

$$f[M_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}; m_{\boldsymbol{p},\boldsymbol{q}}] \sum_{i=1}^{r} \frac{\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - M_{\boldsymbol{p},\boldsymbol{q}}\right)}{H_2(i+q_2)^{s_2}}$$
$$+ \sum_{k=2}^{n-2} f[M_{\boldsymbol{p},\boldsymbol{q}}, M_{\boldsymbol{p},\boldsymbol{q}}; \underbrace{m_{\boldsymbol{p},\boldsymbol{q}}, \ldots, m_{\boldsymbol{p},\boldsymbol{q}}}_{k \text{ times}}] \sum_{i=1}^{r} \frac{\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)^{k-1} \left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - M_{\boldsymbol{p},\boldsymbol{q}}\right)^2}{H_2(i+q_2)^{s_2}}$$
$$\leq \frac{M_{\boldsymbol{p},\boldsymbol{q}} - 1}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} f(m_{\boldsymbol{p},\boldsymbol{q}}) + \frac{1 - m_{\boldsymbol{p},\boldsymbol{q}}}{M_{\boldsymbol{p},\boldsymbol{q}} - m_{\boldsymbol{p},\boldsymbol{q}}} f(M_{\boldsymbol{p},\boldsymbol{q}}) - \tilde{D}_f(\boldsymbol{p},\boldsymbol{q})$$
$$\leq \sum_{k=2}^{n-1} f[M_{\boldsymbol{p},\boldsymbol{q}}; \underbrace{m_{\boldsymbol{p},\boldsymbol{q}}, \ldots, m_{\boldsymbol{p},\boldsymbol{q}}}_{k \text{ times}}] \sum_{i=1}^{r} \frac{\left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - m_{\boldsymbol{p},\boldsymbol{q}}\right)^{k-1} \left(\frac{H_2(i+q_2)^{s_2}}{H_1(i+q_1)^{s_1}} - M_{\boldsymbol{p},\boldsymbol{q}}\right)}{H_2(i+q_2)^{s_2}}.$$

If the function f is n-concave, the inequality signs are reversed.

Remark 8.13 By taking into consideration Remark 8.8 one can see that general results from this section can easily be applied to any of the following divergences: Kullback-Leibler divergence, Hellinger divergence, harmonic divergence or Jeffreys divergence.

8.5 Further generalization of Edmunson-Lah-Ribarič inequality for Zipf-Mandelbrot law

Throughout this section without further noticing when using $[m,M]$ we assume that $-\infty < m < M < \infty$.

Let $r_n(v)$ be defined recursively by

$$r_0(v) = \min\{v, 1-v\}$$
$$r_n(v) = \min\{2r_{n-1}(v), 1-2r_{n-1}(v)\}$$

for $0 \leq v \leq 1$. It has been shown in [5] that

$$r_n(v) = \begin{cases} 2^n v - k + 1 \ , & \frac{k-1}{2^n} \leq v \leq \frac{2k-1}{2^{n+1}}, \\ k - 2^n v \ , & \frac{2k-1}{2^{n+1}} < v \leq \frac{k}{2^n}, \end{cases}$$

for $k = 1, 2, \ldots, 2^n$.

It has been shown (see [5]) that if N is a nonnegative integer and f is convex on $[0,1]$, then

$$(1-v)f(0) + vf(1) \geq f(v) + \sum_{n=0}^{N-1} r_n(v) \sum_{k=1}^{2^n} \Delta_f(n,k)\chi_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right)}(v) \tag{8.71}$$

where

$$\Delta_f(n,k) = f\left(\frac{k-1}{2^n}\right) + f\left(\frac{k}{2^n}\right) - 2f\left(\frac{2k-1}{2^{n+1}}\right),$$

and χ represents the characteristic function of the corresponding interval. If $N = 0$ then sum is zero, that is we have convexity.

In the paper [6] previous relation is extended to hold for an arbitrary interval. Following result is given.

Lemma 8.1 *Let N be a nonnegative integer and let f be convex on $[a,b]$. Then*

$$(1-v)f(a) + vf(b) \geq f((1-v)a + vb) + \sum_{n=0}^{N-1} r_n(v) \sum_{k=1}^{2^n} \Delta_f(a,b,n,k)\chi_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right)}(v) \tag{8.72}$$

where

$$\begin{aligned} \Delta_f(a,b,n,k) \ = \ & f\left(\frac{(2^n-k+1)a + (k-1)b}{2^n}\right) + f\left(\frac{(2^n-k)a + kb}{2^n}\right) \\ & - \ 2f\left(\frac{(2^{n+1}-2k+1)a + (2k-1)b}{2^{n+1}}\right), \end{aligned}$$

and χ represents the characteristic function of the corresponding interval.

Next theorem is main result in paper [32], and it is an improvement of Theorem 8.3.

Theorem 8.30 *Let L satisfy L1, L2, L3 on a nonempty set E and let A be a positive normalized linear functional. If f is a convex function on $[m,M]$ then for all $g \in L$ such that $f(g) \in L$ we have $A(g) \in [m,M]$ and*

$$\begin{aligned} &\frac{M-A(g)}{M-m} f(m) + \frac{A(g)-m}{M-m} f(M) \geq A(f(g)) \\ &+ \sum_{n=0}^{N-1} \sum_{k=1}^{2^n} \Delta_f(m,M,n,k) A\left(\left(r_n \cdot \chi_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right)}\right)\left(\frac{g-m}{M-m}\right)\right) \end{aligned} \tag{8.73}$$

where

$$\begin{aligned} \Delta_f(m,M,n,k) &= f\left(\frac{(2^n-k+1)m+(k-1)M}{2^n}\right) + f\left(\frac{(2^n-k)m+kM}{2^n}\right) \\ &- 2f\left(\frac{\left(2^{n+1}-2k+1\right)m+(2k-1)M}{2^{n+1}}\right), \end{aligned}$$

and χ represents the characteristic function of the corresponding interval.

Remark 8.14 If we write equation (8.73) in the following form

$$\begin{aligned} &\frac{M-A(g)}{M-m} f(m) + \frac{A(g)-m}{M-m} f(M) \geq A(f(g)) \\ &+ \Delta_f(m,M,0,1) \cdot A\left(r_0\left(\frac{g-m}{M-m}\right) \cdot \chi_{\left(\frac{0}{2^0}, \frac{1}{2^0}\right)}\left(\frac{g-m}{M-m}\right)\right) \\ &+ \sum_{n=1}^{N-1} \sum_{k=1}^{2^n} \Delta_f(m,M,n,k) A\left(\left(r_n \cdot \chi_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right)}\right)\left(\frac{g-m}{M-m}\right)\right) \end{aligned}$$

and notice

$$\begin{aligned} &\Delta_f(m,M,0,1) = f(m) + f(M) - 2f\left(\tfrac{m+M}{2}\right) \\ &r_0\left(\tfrac{g-m}{M-m}\right) = \tfrac{1}{2} - \frac{\left|g-\frac{m+M}{2}\right|}{M-m} \\ &\chi_{\left(\frac{0}{2^0}, \frac{1}{2^0}\right)}\left(\tfrac{g-m}{M-m}\right) = \chi_{(0,1)}\left(\tfrac{g-m}{M-m}\right) = 1 \end{aligned}$$

we have that Theorem 8.30 is an improvement of Theorem 8.3.

Corollary 8.18 *Let $\mathbf{p}$ be a nonnegative l-tuple with $P_l = \sum_{i=1}^{l} p_i \neq 0$ and $\mathbf{x} \in [m,M]^l$. If $f\colon [m,M] \to \mathbb{R}$ is a convex function then*

$$\begin{aligned} &\frac{1}{P_l} \sum_{i=1}^{l} p_i f(x_i) \\ &\leq \frac{M-\overline{x}}{M-m} f(m) + \frac{\overline{x}-m}{M-m} f(M) \\ &-\frac{1}{P_l} \sum_{n=0}^{N-1} \sum_{k=1}^{2^n} \sum_{i=1}^{l} \Delta_f(m,M,n,k) p_i \left[\left(r_n \cdot \chi_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right)}\right)\left(\frac{x_i-m}{M-m}\right)\right] \end{aligned} \tag{8.74}$$

$$
\begin{aligned}
&= \frac{M-\overline{x}}{M-m} f(m) + \frac{\overline{x}-m}{M-m} f(M) \\
&-\frac{1}{P_l}\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)p_i\left[\left(2^n\frac{x_i-m}{M-m}-k+1\right)\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{x_i-m}{M-m}\right)\right. \\
&\left.+\left(k-2^n\frac{x_i-m}{M-m}\right)\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{x_i-m}{M-m}\right)\right]
\end{aligned}
$$

where $\overline{x} = \frac{1}{P_l}\sum_{i=1}^{l} p_i x_i$.

Proof. If we consider $E = [m,M]$, $L = \mathbb{R}^{[m,M]}$, $g = id_E$, $A(f) = \frac{1}{P_l}\sum_{i=1}^{l} p_i f(x_i)$ in Theorem 8.30, then inequality (8.73) becomes (8.74). □

Using previous result we get some improvements of some results from Section 8.3. First we give an improvement of Theorem 8.17.

Theorem 8.31 *Let $[m,M] \subset \mathbb{R}$ be an interval and let $f\colon [m,M] \to \mathbb{R}$ be a function. Let $\boldsymbol{p} = (p_1,...,p_l)$ be an l-tuple of real numbers and $\boldsymbol{q} = (q_1,...,q_l)$ be an l-tuple of nonnegative real numbers such that $p_i/q_i \in [m,M]$ for every $i = 1,\ldots,l$. If the function f is convex, we have*

$$
\begin{aligned}
\hat{D}_f(\boldsymbol{p},\boldsymbol{q}) &\le \frac{MQ_l-P_l}{M-m}f(m)+\frac{P_l-mQ_l}{M-m}f(M) \\
&-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)q_i\left[\left(r_n\cdot\chi_{\left(\frac{k-1}{2^n},\frac{k}{2^n}\right)}\right)\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right] \\
&= \frac{MQ_l-P_l}{M-m}f(m)+\frac{P_l-mQ_l}{M-m}f(M) \\
&-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)\frac{1}{M-m}\cdot \\
&\cdot\left[(2^n(p_i-mq_i)-q_i(M-m)(k-1))\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right. \\
&\left.+(q_i(M-m)k-2^n(p_i-mq_i))\,\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right]
\end{aligned}
$$

where $P_l = \sum_{i=1}^{l} p_i$ and $Q_l = \sum_{i=1}^{l} q_i$. If the function f is concave, then the inequality sign is reversed.

Proof. Let $f\colon [m,M] \to \mathbb{R}$ be a convex function. For an l-tuple of real numbers $\boldsymbol{x} = (x_1,...,x_l)$ and an l-tuple of nonnegative numbers $\boldsymbol{p} = (p_1,...,p_l)$ from Corollary 8.18 we have

$$
\begin{aligned}
&\frac{1}{P_l}\sum_{i=1}^{l} p_i f(x_i) \\
\leq\ & \frac{M-\overline{x}}{M-m} f(m) + \frac{\overline{x}-m}{M-m} f(M) \\
&-\frac{1}{P_l}\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)p_i\left[\left(r_n\cdot\chi_{\left(\frac{k-1}{2^n},\frac{k}{2^n}\right)}\right)\left(\frac{x_i-m}{M-m}\right)\right] \\
=\ & \frac{M-\overline{x}}{M-m} f(m) + \frac{\overline{x}-m}{M-m} f(M) \\
&-\frac{1}{P_l}\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)p_i\left[\left(2^n\frac{x_i-m}{M-m}-k+1\right)\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{x_i-m}{M-m}\right)\right. \\
&\left.+\left(k-2^n\frac{x_i-m}{M-m}\right)\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{x_i-m}{M-m}\right)\right]
\end{aligned}
$$

where $\overline{x}=\frac{1}{P_l}\sum_{i=1}^{l} p_i x_i$. Since $\boldsymbol{q}=(q_1,\ldots,q_l)$ are nonnegative real numbers, we can put

$$
p_i = q_i \ \text{ and } \ x_i = \frac{p_i}{q_i}
$$

in previous inequality and get

$$
\begin{aligned}
&\frac{1}{Q_l}\sum_{i=1}^{l} q_i f\left(\frac{p_i}{q_i}\right) \leq \frac{M-\frac{1}{Q_l}\sum_{i=1}^{l} q_i\frac{p_i}{q_i}}{M-m} f(m) + \frac{\frac{1}{Q_l}\sum_{i=1}^{l} q_i\frac{p_i}{q_i}-m}{M-m} f(M) \\
&-\frac{1}{Q_l}\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)q_i\left[r_n\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\cdot\chi_{\left(\frac{k-1}{2^n},\frac{k}{2^n}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right] \\
=\ & \frac{M-\frac{1}{Q_l}\sum_{i=1}^{l} q_i\frac{p_i}{q_i}}{M-m} f(m) + \frac{\frac{1}{Q_l}\sum_{i=1}^{l} q_i\frac{p_i}{q_i}-m}{M-m} f(M) \\
&-\frac{1}{Q_l}\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)\frac{1}{M-m}\cdot \\
&\cdot\left[(2^n(p_i-mq_i)-q_i(M-m)(k-1))\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right. \\
&\left.+(q_i(M-m)k-2^n(p_i-mq_i))\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right]
\end{aligned}
$$

and after multiplying by Q_l we get the result. □

Remark 8.15 Analogously as in Remark 8.14 we see that previous result is improvement Theorem 8.17.

Remark 8.16 If in the previous theorem we take $\boldsymbol{p}$ and $\boldsymbol{q}$ to be probability distributions, we directly get following result for the Csiszár f-divergence functional, which is an improvement of Remark 8.2.

$$\begin{aligned}
D_f(\boldsymbol{p},\boldsymbol{q}) &\le \frac{M-1}{M-m}f(m)+\frac{1-m}{M-m}f(M)\\
&-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)q_i\left[r_n\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\cdot\chi_{\left(\frac{k-1}{2^n},\frac{k}{2^n}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right]\\
&= \frac{M-1}{M-m}f(m)+\frac{1-m}{M-m}f(M)\\
&-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)\frac{1}{M-m}\cdot\\
&\cdot\left[(2^n(p_i-mq_i)-q_i(M-m)(k-1))\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right.\\
&\left.+(q_i(M-m)k-2^n(p_i-mq_i))\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right]
\end{aligned}$$

Next result provides us with improvement of the bounds for the Kullback-Leibler divergence of two probability distributions, that is result from Corollary 8.1.

Corollary 8.19 *Let $[m,M]\subset\mathbb{R}$ be an interval and let us assume that the base of the logarithm is greater than* 1.

- *Let $\boldsymbol{p}=(p_1,\dots,p_n)$ and $\boldsymbol{q}=(q_1,\dots,q_l)$ be l-tuples of nonnegative real numbers such that $\frac{p_i}{q_i}\in[m,M]$ for every $i=1,\dots,l$. Then*

$$\begin{aligned}
\sum_{i=1}^{l}p_i\log\left(\frac{p_i}{q_i}\right) &\le Q_l\frac{Mm}{M-m}\log\left(\frac{m}{M}\right)+\frac{P_l}{M-m}\log\left(\frac{M^M}{m^m}\right)\\
&-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_{log}(m,M,n,k)q_i\left[\left(r_n\cdot\chi_{\left(\frac{k-1}{2^n},\frac{k}{2^n}\right)}\right)\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right]\\
&= Q_l\frac{Mm}{M-m}\log\left(\frac{m}{M}\right)+\frac{P_l}{M-m}\log\left(\frac{M^M}{m^m}\right)\\
&-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_{log}(m,M,n,k)\frac{1}{M-m}\cdot\\
&\cdot\left[(2^n(p_i-mq_i)-q_i(M-m)(k-1))\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right.\\
&\left.+(q_i(M-m)k-2^n(p_i-mq_i))\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right]
\end{aligned}$$

- *Let $\boldsymbol{p} = (p_1,...,p_n)$ and $\boldsymbol{q} = (q_1,...,q_n) \in \mathbb{P}$ be probability distributions such that $m \leq p_i/q_i \leq M$ holds for every $i = 1,...,n$. Then*

$$
\begin{aligned}
D_{KL}(\boldsymbol{p},\boldsymbol{q}) &\leq \frac{Mm}{M-m}\log\left(\frac{m}{M}\right) + \frac{1}{M-m}\log\left(\frac{M^M}{m^m}\right) \\
&- \sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_{log}(m,M,n,k)q_i\left[r_n\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\cdot\chi_{\left(\frac{k-1}{2^n},\frac{k}{2^n}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right] \\
= &\frac{Mm}{M-m}\log\left(\frac{m}{M}\right) + \frac{1}{M-m}\log\left(\frac{M^M}{m^m}\right) \\
&- \sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_{log}(m,M,n,k)\frac{1}{M-m}\cdot \\
&\cdot\left[(2^n(p_i-mq_i)-q_i(M-m)(k-1))\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right. \\
&\left.+(q_i(M-m)k-2^n(p_i-mq_i))\,\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{\frac{p_i}{q_i}-m}{M-m}\right)\right]
\end{aligned}
$$

If the base of the logarithm is less than 1*, the inequality sign in the inequalities above is reversed.*

Proof. Let $\boldsymbol{p} = (p_1,...,p_n)$ and $\boldsymbol{q} = (q_1,...,q_n)$ be an n-tuples of nonnegative real numbers. Since the function $t \mapsto t\log t$ is convex when the base of the logarithm is greater than 1, first inequality follows from Theorem 8.31 by setting $f(t) = t\log t$.

Second inequality is a special case of the first inequality for probability distributions $\boldsymbol{p}$ and $\boldsymbol{q}$. □

Remark 8.17 Analogously as in Remark 8.14 we see that previous result is improvement Corollary 8.1.

Last results in this section will be about Zipf-Mandelbrot law.

We will denote in this section parameters in Zipf-Mandelbrot law as l,t_1,s_1 because of the previous results.
If we define **q** as a Zipf-Mandelbrot law l-tuple, we have

$$
q_i = \frac{1}{(i+t_2)^{s_2}H_{l,s_2,t_2}}, i = 1,\ldots,l
$$

where

$$
H_{l,s_2,t_2} = \sum_{i=1}^{l}\frac{1}{(k+t_2)^{s_2}}
$$

and Csiszar functional becomes

$$
\hat{D}_f(\mathbf{p},i,l,s_2,t_2) = \sum_{i=1}^{l}\frac{1}{(i+t_2)^{s_2}H_{l,s_2,t_2}}f\left(p_i(i+t_2)^{s_2}H_{l,s_2,t_2}\right),
$$

where $f\colon I \to \mathbb{R}, I \subseteq \mathbb{R}$, and the parameters $l \in \mathbb{N}, s_2 > 0, t_2 \geq 0$ are such that $p_i(i+t_2)^{s_2} H_{l,s_2,t_2} \in I, i = 1, \ldots, l$.

If $\mathbf{p}$ and $\mathbf{q}$ are both defined as Zipf-Mandelbrot law l-tuples, then Csiszar functional becomes

$$\hat{D}_f(i,l,s_1,s_2,t_1,t_2) = \sum_{i=1}^{l} \frac{1}{(i+t_2)^{s_2} H_{l,s_2,t_2}} f\left(\frac{(i+t_2)^{s_2} H_{l,s_2,t_2}}{(i+t_1)^{s_1} H_{l,s_1,t_1}}\right),$$

where $f\colon I \to \mathbb{R}, I \subseteq \mathbb{R}$, and the parameters $l \in \mathbb{N}, s_1, s_2 > 0, t_1, t_2 \geq 0$ are such that $\frac{(i+t_2)^{s_2} H_{l,s_2,t_2}}{(i+t_1)^{s_1} H_{l,s_1,t_1}} \in I, i = 1, \ldots, l$.

Since the minimal value for q_i is $\min\{q_i\} = \frac{1}{(l+t_2)^{s_2} H_{l,s_2,t_2}}$, then from Theorem 8.31 we have following result.

Corollary 8.20 *Let $\mathbf{p} = (\mathbf{p_1}, \ldots, \mathbf{p_l})$ be an l-tuple of real numbers with $P_l = \sum_{i=1}^{l} p_i$. Suppose $I \subseteq \mathbb{R}$ is an interval, $l \in \mathbb{N}$ and $s_2 > 0, t_2 \geq 0$ are such that $p_i(i+t_2)^{s_2} H_{l,s_2,t_2} \in I, i = 1, \ldots, l$. If $f\colon I \to \mathbb{R}$ is a convex function, then*

$$\begin{aligned}
&\hat{D}_f(\boldsymbol{p},i,l,s_2,t_2) \leq \frac{M-P_l}{M-m} f(m) + \frac{P_l-m}{M-m} f(M) \\
&\quad - \sum_{n=0}^{N-1} \sum_{k=1}^{2^n} \sum_{i=1}^{l} \Delta_f(m,M,n,k) \frac{1}{(l+t_2)^{s_2} H_{l,s_2,t_2}} \left[\left(r_n \cdot \chi_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right)} \right) \left(\frac{p_i(i+t_2)^{s_2} H_{M,s_2,t_2} - m}{M-m} \right) \right] \\
&\leq \frac{M-P_l}{M-m} f(m) + \frac{P_l-m}{M-m} f(M) - \sum_{n=0}^{N-1} \sum_{k=1}^{2^n} \sum_{i=1}^{l} \Delta_f(m,M,n,k) \frac{1}{M-m} \\
&\quad \left[(2^n(p_i - m\min\{q_i\}) - \min\{q_i\}(M-m)(k-1)) \cdot \chi_{\left(\frac{k-1}{2^n}, \frac{2k-1}{2^{n+1}}\right)} \left(\frac{p_i(i+t_2)^{s_2} H_{M,s_2,t_2} - m}{M-m} \right) \right. \\
&\quad \left. + (\min\{q_i\}(M-m)k - 2^n(p_i - m\min\{q_i\})) \chi_{\left(\frac{2k-1}{2^{n+1}}, \frac{k}{2^n}\right)} \left(\frac{p_i(i+t_2)^{s_2} H_{M,s_2,t_2} - m}{M-m} \right) \right]
\end{aligned}$$

Proof. Follows easily from Theorem 8.31. □

Now let's denote

$$\begin{aligned}
&H_{l,s_1,t_1} = H_1, H_{l,s_2,t_2} = H_2, \\
&m_{\mathbf{p},\mathbf{q}} := \min\left\{\frac{p_i}{q_i}\right\} = \frac{H_2}{H_1} \min\left\{\frac{(i+q_2)^{s_2}}{(i+q_1)^{s_1}}\right\}.
\end{aligned}$$

Corollary 8.21 *Let $I \subseteq \mathbb{R}$ be an interval and suppose $N \in \mathbb{N}, s_1, s_2 > 0, q_1, q_2 \geq 0$ are such that $\frac{(i+t_2)^{s_2} H_{l,s_2,t_2}}{(i+t_1)^{s_1} H_{l,s_1,t_1}} \in I, i = 1, \ldots, l$.*
If $f\colon I \to \mathbb{R}$ is a convex function, then

$$\begin{aligned}
&\hat{D}_f(i,l,s_1,s_2,t_1,t_2) \leq \frac{M-1}{M-m} f(m) + \frac{1-m}{M-m} f(M) \\
&\quad - \sum_{n=0}^{N-1} \sum_{k=1}^{2^n} \sum_{i=1}^{l} \Delta_f(m,M,n,k) \frac{1}{(l+t_2)^{s_2} H_{l,s_2,t_2}} \left[\left(r_n \cdot \chi_{\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right)} \right) \left(\frac{\frac{(i+t_2)^{s_2} H_{l,s_2,t_2}}{(i+t_1)^{s_1} H_{l,s_1,t_1}} - m}{M-m} \right) \right]
\end{aligned}$$

$$\leq \frac{M-1}{M-m}f(m)+\frac{1-m}{M-m}f(M)-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_f(m,M,n,k)\frac{1}{(l+t_2)^{s_2}H_{l,s_2,t_2}}\cdot$$
$$\cdot\left[\left(2^n\left(\frac{m_{\mathbf{p},\mathbf{q}}-m}{M-m}\right)-k+1\right)\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{\frac{(i+t_2)^{s_2}H_{l,s_2,t_2}}{(i+t_1)^{s_1}H_{l,s_1,t_1}}-m}{M-m}\right)\right.$$
$$\left.+\left(k-2^n\left(\frac{m_{\mathbf{p},\mathbf{q}}-m}{M-m}\right)\right)\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{\frac{(i+t_2)^{s_2}H_{l,s_2,t_2}}{(i+t_1)^{s_1}H_{l,s_1,t_1}}-m}{M-m}\right)\right].$$

Proof. Follows easily from Theorem 8.31. □

We denote Kullback-Leibler divergence for **p** and **q** both defined as Zipf-Mandelbrot law l-tuples as $D_{KL}(i,l,s_1,s_2,t_1,t_2)$.

Corollary 8.22 *Let $l \in \mathbb{N}$ and $s_1,s_2 > 0, t_1,t_2 \geq 0$.*
If the logarithm base is greater than 1, then

$$D_{KL}(\boldsymbol{p},\boldsymbol{q}) \leq \frac{Mm}{M-m}\log\left(\frac{m}{M}\right)+\frac{1}{M-m}\log\left(\frac{M^M}{m^m}\right)$$
$$-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_{log}(m,M,n,k)\frac{1}{(l+t_2)^{s_2}H_{l,s_2,t_2}}\left[\left(r_n\cdot\chi_{\left(\frac{k-1}{2^n},\frac{k}{2^n}\right)}\right)\left(\frac{\frac{(i+t_2)^{s_2}H_{l,s_2,t_2}}{(i+t_1)^{s_1}H_{l,s_1,t_1}}-m}{M-m}\right)\right]$$
$$\leq \frac{Mm}{M-m}\log\left(\frac{m}{M}\right)+\frac{1}{M-m}\log\left(\frac{M^M}{m^m}\right)$$
$$-\sum_{n=0}^{N-1}\sum_{k=1}^{2^n}\sum_{i=1}^{l}\Delta_{log}(m,M,n,k)\frac{1}{(l+t_2)^{s_2}H_{l,s_2,t_2}}\cdot$$
$$\cdot\left[\left(2^n\left(\frac{m_{\mathbf{p},\mathbf{q}}-m}{M-m}\right)-k+1\right)\cdot\chi_{\left(\frac{k-1}{2^n},\frac{2k-1}{2^{n+1}}\right)}\left(\frac{\frac{(i+t_2)^{s_2}H_{l,s_2,t_2}}{(i+t_1)^{s_1}H_{l,s_1,t_1}}-m}{M-m}\right)\right.$$
$$\left.+\left(k-2^n\left(\frac{m_{\mathbf{p},\mathbf{q}}-m}{M-m}\right)\right)\chi_{\left(\frac{2k-1}{2^{n+1}},\frac{k}{2^n}\right)}\left(\frac{\frac{(i+t_2)^{s_2}H_{l,s_2,t_2}}{(i+t_1)^{s_1}H_{l,s_1,t_1}}-m}{M-m}\right)\right].$$

If the base of the logarithm is less than 1*, the inequality sign in the inequalities above is reversed.*

Proof. Follows easily from Corollary 8.19. □

Bibliography

[1] R. P. Agarwal, P. J. Y. Wong, *Error Inequalities in Polynomial Interpolation and Their Applications*, Kluwer Academic Publishers, Dordrecht, Boston, London, 1993.

[2] M. Ben Bassat, *f-entropies, probability of error, and feature selection*, Inform. Contr., **39**, (1978), 227–242.

[3] P. R. Beesack, J. E. Pečarić, *On the Jessen's inequality for convex functions*, J. Math. Anal. **110**, (1985), 536–552.

[4] C. H. Chen, *Statistical Pattern Recognition*, Rochelle Park, NJ: Hayden Book Co., 1973.

[5] D. Choi, M. Krnić, J. Pečarić, *Improved Jensen-type inequalities via linear interpolation and applications*, Math. Inequal. 11 (2017), 301–322.

[6] D. Choi, M. Krnić, J. Pečarić, *More accurate classes of Jensen-type inequalities for convex and operator convex functions*, submitted for publication.

[7] C. K. Chow, C. N. Liu, *Approximating discrete probability distributions with dependence trees*, IEEE Trans. Inform. Theory, **14** (3), (1968), 462–467.

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

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

[10] V. Diodato, *Dictionary of Bibliometrics*, Haworth Press, New York: 1994.

[11] S. S. Dragomir, *Other inequalities for Csiszár divergence and applications*, RGMIA.

[12] L. Egghe, R. Rousseau, *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science*, Elsevier Science Publishers. New York: 1990.

[13] A. Gelbukh, G. Sidorov *Zipf and Heaps Laws Coefficients Depend on Language*, Proc. CICLing-2001, Conference on Intelligent Text Processing and Computational Linguistics, Mexico City.

[14] D. V. Gokhale, S. Kullback, *Information in Contingency Tables*, New York: Marcel Dekker, 1978.

[15] L. Horvath, G. Pečarić, J. Pečarić, *Estimations of f- and Rènyi divergences by using a cyclic refinement of the Jensen's inequality*, Bull. Malays. Math. Sci. Soc. (2017). doi:10.1007/s40840-017-0526-4.

[16] E. Issacson, H. B. Keller, *Analysis of Numerical methods*, Dover Publications Inc., New York: 1966.

[17] B. Jessen, (1931). *Bemaerkinger om konvekse Funktioner og Uligheder imellem Middelvaerdier I*, Mat. Tidsskrift, B, 17–28.

[18] T. Kailath, *The divergence and Bhattacharyya distance measures in signal selection*, IEEE Transuctions Commun. Technol., **15** (1), (1967), 52–60.

[19] M. Klaričić Bakula, J. E. Pečarić, J. Perić *On the converse Jensen inequality*, Appl. Math. Comput. **218**, (2012).

[20] M. Krnić, R. Mikić, J. Pečarić, *Strengthened converses of the Jensen and Edmundson-Lah-Ribari inequalities*, Adv. Oper. Theory **1** (1), (2016), 104–122.

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

[22] J. Lin, S. K. M. Wong, *Approximation of discrete probability distributions based on a new divergence measure*, Congressus Numerantiitm, **61**, (1988), 75–80.

[23] B. Manaris, D. Vaughan, C. S. Wagner, J. Romero, R. B. Davis, *Evolutionary Music and the Zipf-Mandelbrot Law: Developing Fitness Functions for Pleasant Music*, Proceedings of 1st European Workshop on Evolutionary Music and Art (EvoMUSART2003), 522–534.

[24] R. Mikić, Đ. Pečarić, J. Pečarić, *Inequalities of the Jensen and Edmundson-Lah-Ribarič type for 3-convex functions with applications*, J. Math. Inequal., **12** (3), (2018), 677–692.

[25] R. Mikić, Đ. Pečarić, J. Pečarić, *Converses of the Edmundson-Lah-Ribarič inequality for generalized Csiszár divergence with applications to Zipf-Mandelbrot law*, An. Univ. Craiova Ser. Mat. Inform. **45** (2), (2018), 243–257.

[26] R. Mikić, Đ. Pečarić, J. Pečarić, *Inequalities of the Jensen and Edmundson-Lah-Ribarič type for positive linear functionals with applications*, Moscow Math. J. **18** (4), (2018), 739–753.

[27] R. Mikić, Đ. Pečarić, J. Pečarić, *Some inequalities of the Edmundson-Lah-Ribarič type for 3-convex functions with applications*, submitted for publication.

[28] R. Mikić, Đ. Pečarić, J. Pečarić, *Inequalities of the Edmundson-Lah-Ribarič type for n-convex functions with applications*, arXiv:1809.08813.

[29] M. A. Montemurro, *Beyond the Zipf-Mandelbrot law in quantitative linguistics*, (2001) URL: arXiv:cond-mat/0104066v2 (22.2.2017.)

[30] D. Mouillot, A. Lepretre, *Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity*, Environmental Monitoring and Assessment. Springer. **63** (2), (2000), 279–295.

[31] M. E. J. Newman, *Power laws, Pareto distributions and Zipfs law*, URL arxiv.org/pdf/cond-mat/0412004.pdf?origin=publication_detail (6.4.2014.)

[32] J. E. Pečarić, J. Perić, *New improvement of the converse Jensen inequality*, Math. Inequal. Appl., 21 (2018), 1, 217–234.

[33] J. E. Pečarić, F. Proschan, Y. L. Tong, *Convex functions, Partial orderings and statistical applications*, Academic Press Inc., San Diego 1992.

[34] Z. K. Silagadze, *Citations and the Zipf-Mandelbrot Law* Complex Systems, 1997 (11), pp: 487–499.

Chapter 9

On Sherman's inequality with applications in information theory

Ana Barbir, Slavica Ivelić Bradanović, Đilda Pečarić and Josip Pečarić

Abstract. In this paper we proved a converse to Sherman's inequality. Using the concept of f-divergence we obtained a converse to the Csiszár-Körner inequality and some inequalities for the well-known entropies. We also established a new lower and upper bounds for Sherman's inequality as well as for f-divergence functional using some basic convexity facts. As special cases and corollaries of obtained bounds we establishe lower and upper bounds for Shannon's entropy and relative entropy also known as the Kullback-Leibler divergence. We also introduced a new entropy by applying the Zipf-Mandelbrot law and derived some related inequalities.

9.1 Introduction and preliminaries

In the space $\mathbb{R}^n$, in which the order is not defined, the term majorization is introduced to compare and detect potential links between vectors.

For two vectors $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, \ldots, y_n)$ from $\mathbb{R}^n$, we say that $\mathbf{x}$ *majorizes*

$\mathbf{y}$ or $\mathbf{y}$ is *majorized* by $\mathbf{x}$ and write

$$\mathbf{y} \prec \mathbf{x}$$

if

$$\sum_{i=1}^{k} y_{[i]} \leq \sum_{i=1}^{k} x_{[i]}, \quad k = 1, \ldots, m-1, \tag{9.1}$$

and

$$\sum_{i=1}^{m} y_i = \sum_{i=1}^{m} x_i.$$

Here

$$x_{[1]} \geq x_{[2]} \geq \ldots \geq x_{[m]}, \quad y_{[1]} \geq y_{[2]} \geq \ldots \geq y_{[m]},$$

are their ordered components.

Note that (9.1) is equivalent to

$$\sum_{i=m-k+1}^{m} y_{(i)} \leq \sum_{i=m-k+1}^{m} x_{(i)}, \quad k = 1, \ldots, m-1,$$

where

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(m)}, \quad y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(m)}$$

This definition defines relation which is reflexive and transitive but it is not antisymmetric (see [23]). Hence, it is a preordering not a partial ordering on $\mathbb{R}^n$. It it important to remember that two vectors may not have any majorization relationship, meaning that there are vectors that can not be compared in the sense of the given definition. Let's assume for example $\boldsymbol{x} = (0.6, 0.6, 0.2)$ and $\boldsymbol{y} = (0.5, 0.4, 0.1)$ vectors for which no $\mathbf{y} \prec \mathbf{x}$ or $\mathbf{x} \prec \mathbf{y}$ are valid.

Issai Schur ([30], [31]) first systematically studied functions that preserve the order of majorization and called them convex functions. Such functions are now called Schur convex functions and convex functions are often referred to convex functions in the sense of Jensen. Each function that is convex and symmetric is also Schur-convex. The opposite implication is not valid. However, all Schur convex functions are symmetric. The concept of majorization, together with the concept of Schur convexity, provides an important characterization of convex functions and is a powerful and useful mathematical tool that has wide application in many applied sciences. Many key ideas related to the concept of majorization are presented in the monograph [2]. However, this monograph attracted only a relatively small number of researchers in terms of problems related to the concept of majorization. The publication of monograph [23] contributes a great interest in the potential application of the concept of majorization and Schur convexity in various scientific areas. It provides a systematic overview of past results from the field of mathematical inequalities with a special emphasis on majorization for the first time, and has been complemented by many new results. One of the important results from the mentioned monographs is the known majorization theorem that gives the relationship between one-dimensional convex functions and n-dimensional Schur convex functions. Majorization theorem can also be expressed in the form of the inequality

$$\sum_{i=1}^{n} f(y_i) \leq \sum_{i=1}^{n} f(x_i), \tag{9.2}$$

which is valid for every function $f:[\alpha,\beta]\to\mathbb{R}$, continuous and convex on some interval $[\alpha,\beta]\subset\mathbb{R}$, where vectors $\mathbf{x},\mathbf{y}\in[\alpha,\beta]^n$ are such that $\mathbf{y}\prec\mathbf{x}$.

A natural problem of interest is extension of notation from n-tuples (vectors) to $n\times m$ matrices $\mathbf{A}=(a_{ij})\in\mathscr{M}_{nm}(\mathbb{R})$. In order that, we introduce the notion of row stochastic and double stochastic matrices.

A matrix $\mathbf{A}=(a_{ij})\in\mathscr{M}_{nm}(\mathbb{R})$ is called *column stochastic* if all of its entries are greater or equal to zero, i.e. $a_{ij}\geq 0$ for $i=1,\dots,n$, $j=1,\dots,m$ and the sum of the entries in each column is equal to 1, i.e. $\sum\limits_{i=1}^{n}a_{ij}=1$ for $j=1,\dots,m$.

A square matrix $\mathbf{A}=(a_{ij})\in\mathscr{M}_m(\mathbb{R})$ is called *double stochastic* if both $\mathbf{A}$ and its transpose $\mathbf{A}^T=(a_{ji})$ are row stochastic. In other words, $\mathbf{A}=(a_{ij})\in\mathscr{M}_m(\mathbb{R})$ is called *double stochastic* if all of its entries are greater or equal to zero, i.e. $a_{ij}\geq 0$ for $i,j=1,\dots,m$, and the sum of the entries in each column and each row is equal to 1, i.e. $\sum\limits_{i=1}^{m}a_{ij}=1$ for $j=1,\dots,m$ and $\sum\limits_{j=1}^{m}a_{ij}=1$ for $i=1,\dots,m$.

It is well known that for $\mathbf{x},\mathbf{y}\in\mathbb{R}^m$ is valid

$$\mathbf{y}\prec\mathbf{x}\quad\text{if and only if}\quad\mathbf{y}=\mathbf{x}\mathbf{A}$$

for some double stochastic matrix $\mathbf{A}\in\mathscr{M}_m(\mathbb{R})$.

S. Sherman [33], considering the weighting concept of majorisation between vectors $\mathbf{x}=(x_1,\dots,x_n)\in[\alpha,\beta]^n$ and $\mathbf{y}=(y_1,\dots,y_m)\in[\alpha,\beta]^m$, with nonegative weights $\mathbf{a}=(a_1,\dots,a_n)\in[0,\infty)^n$, and $\mathbf{b}=(b_1,\dots,b_m)\in[0,\infty)^m$, taking into account the relation

$$\mathbf{y}=\mathbf{x}S\text{ and }\mathbf{a}=\mathbf{b}S^{\mathsf{T}},\tag{9.3}$$

where $S=(s_{ij})\in\mathscr{M}_{nm}(\mathbb{R})$ is some column stochastic matrix and S^{T} a transpose matrix of S, proved that more general inequality

$$\sum_{j=1}^{m}b_jf(y_j)\leq\sum_{i=1}^{n}a_if(x_i)\tag{9.4}$$

is valid for every function $f:[\alpha,\beta]\to\mathbb{R}$ convex on some interval $[\alpha,\beta]\subset\mathbb{R}$.

Note that (9.3) can be written as

$$\begin{aligned}\mathbf{y}=\mathbf{x}S,&\quad\left(y_j=\sum_{i=1}^{n}x_is_{ij},\ \ j=1,\dots,m\right),\\ \mathbf{a}=\mathbf{b}S^{\mathsf{T}},&\quad\left(a_i=\sum_{j=1}^{m}b_js_{ij},\ \ i=1,\dots,n\right).\end{aligned}\tag{9.5}$$

Inequality (9.1) follows from Sherman's inequality (9.4) as a simple consequence. Moreover, Sherman's inequality (9.4) reduces to Jensen's inequality

$$f\left(\sum_{i=1}^{n}a_ix_i\right)\leq\sum_{i=1}^{n}a_if(x_i),\tag{9.6}$$

where $\boldsymbol{x} = (x_1,\ldots,x_n)$ is any n-tuple in $[\alpha,\beta]^n$ and $\boldsymbol{a} = (a_1,\ldots,a_n) \in [0,\infty)^n$ such that $\sum_{i=1}^n a_i = 1$. It is obviously true, by choosing $m = 1$ and setting $\mathbf{b} = [1]$.

On the other hand, Sherman's inequality (9.4) is an easy consequence of Jensen's inequality (9.6), i.e. under the assumptions (9.3) we have

$$\begin{aligned}\sum_{j=1}^m b_j f(y_j) &= \sum_{j=1}^m b_j f\left(\sum_{i=1}^n x_i s i j\right)\\ &\leqslant \sum_{j=1}^m b_j \sum_{i=1}^n s_{ij} f(x_i)\\ &= \sum_{i=1}^n \left(\sum_{j=1}^m b_j s_{ij}\right) f(x_i) = \sum_{i=1}^n a_i f(x_i).\end{aligned}$$

Considering the difference between the right and left side of Jensen's inequality (9.6), we define the normalized Jensen functional for a convex function as follows

$$J_n(f,\mathbf{x},\mathbf{a}) = \sum_{i=1}^n a_i f(x_i) - f\left(\sum_{i=1}^n a_i x_i\right) \geq 0.$$

Dragomir [34] obtained the lower and upper bound for the normalized Jensen functional stated in the next theorem.

Theorem 9.1 *Let* $[\alpha,\beta] \subset \mathbb{R}$, $\mathbf{x} = (x_1,\ldots,x_n) \in [\alpha,\beta]^n$ *and* $\mathbf{a} = (p_1,\ldots,a_n) \in [0,\infty)^n$ *with* $\sum_{i=1}^n a_i = 1$. *Then for every convex function* $f : [\alpha,\beta] \to \mathbb{R}$, *the inequality*

$$0 \leqslant \min_{1\leqslant i\leqslant n}\{a_i\} S_f(\mathbf{x}) \leqslant J_n(f,\mathbf{x},\mathbf{a}) \leqslant \max_{1\leqslant i\leqslant n}\{a_i\} S_f(\mathbf{x}) \tag{9.7}$$

holds, where $S_f(\mathbf{x})$ *is defined by*

$$S_f(\mathbf{x}) = \sum_{i=1}^n f(x_i) - nf\left(\frac{1}{n}\sum_{i=1}^n x_i\right). \tag{9.8}$$

We close this introduction with one more inequality closely connected to Jensen's inequality (9.6) known as the Lah-Ribarič inequality

$$\sum_{i=1}^n a_i f(x_i) \leq \frac{\beta - \overline{x}}{\beta - \alpha} f(\alpha) + \frac{\overline{x} - \alpha}{\beta - \alpha} f(\beta) \tag{9.9}$$

which holds for every function $f : [\alpha,\beta] \to \mathbb{R}$ convex on $[\alpha,\beta] \subset \mathbb{R}$, where $\mathbf{x} = (x_1,\ldots,x_n) \in [\alpha,\beta]^n$, $\mathbf{a} = (a_1,\ldots,a_n) \in [0,\infty)^n$ with $\sum_{i=1}^n a_i = 1$ and $\overline{x} = \sum_{i=1}^n a_i x_i$ (see [21], [28]).

The contents of this chapter corresponds for the most part to the contents of the papers [11] and [14].

9.2 The concept of Csiszár f-divergence functional

Csiszár [5] introduced the concept of f-divergence functional

$$C_f(\mathbf{q},\mathbf{p}) = \sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right) \tag{9.10}$$

for a convex function $f:(0,\infty)\to\mathbb{R}$ and $\mathbf{p}=(p_1,\ldots,p_n)\in(0,\infty)^n$, $\mathbf{q}=(q_1,\ldots,q_n)\in(0,\infty)^n$.

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 $f:(0,\infty)\to\mathbb{R}$ is defined by

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right), \tag{9.11}$$

with weights $r_1,\ldots,r_n\geq 0$.

Remark 9.1 Notice that $C_f(\mathbf{q},\mathbf{p};\mathbf{e})=C_f(\mathbf{q},\mathbf{p})$ for $\mathbf{e}=(1,\ldots,1)\in\mathbb{R}^n$.

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

$$\sum_{i=1}^{n} p_i f\left(\frac{\sum_{i=1}^{n} q_i}{\sum_{i=1}^{n} p_i}\right) \leqslant C_f(\mathbf{q},\mathbf{p}) \tag{9.12}$$

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

$$0\leq C_f(\mathbf{q},\mathbf{p}). \tag{9.13}$$

In particular, if $\mathbf{p}$ and $\mathbf{q}$ are two positive probability distribution, i.e. $\mathbf{p}=(p_1,\ldots,p_n)\in(0,\infty)^n$ and $\mathbf{q}=(q_1,\ldots,q_n)\in(0,\infty)^n$ with $\sum_{i=1}^{n} p_i=\sum_{i=1}^{n} q_i=1$, then the inequality (9.13) holds for every convex and normalized function $f:(0,\infty)\to\mathbb{R}$. These results are easy consequences of Jensen's inequality (9.6).

We may consider Csiszár's f-divergence functional as generalized measure of information on the set of probability distribution. It is not metric. However, it satisfies the triangle inequality but it is not symmetric. The general aspect of the Csiszár f-divergence functional (9.10) can be interpreted as a series of well-known entropies, divergencies and distances for suitable choices of the kernel f. Entropies quantify the diversity, uncertainty and randomness of a system. The idea has been widely employed in different scientic fields among which we point out mathematical statistics and specially information theory (see [4], [7], [17], [18], [19], [20], [29], [32]). Information theory is a mathematical theory of learning with deep connections with topics as diverse as artificial intelligence, statistical physics, and biological evolution.

In the examples below, for suitable choices of the kernel f, we obtain some of the best known distance functions.

Example 9.1 As a special case from the Csiszáre f-divergence, choosing the convex mapping $f(t) = \ln\frac{1}{t} = -\ln t$, $t > 0$, we get the Shannon entropy defined by

$$H(\mathbf{p}) = \sum_{i=1}^{n} p_i \ln\left(\frac{1}{p_i}\right) = -C_f(\mathbf{e},\mathbf{p}). \tag{9.14}$$

This a statistical concept of entropy is introduced by Shannon [32] in the theory of communication and transmission of information as a measure of information.
We also consider the concept of weighted Shannon's entropy defined by

$$H(\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i \ln\left(\frac{1}{p_i}\right) = -C_f(\mathbf{e},\mathbf{p};\mathbf{r}), \tag{9.15}$$

introduced by Belis and Guia**c**su, motivated by various communication and transmission problems, taking into account probabilities and some qualitative characteristic of events. If we ignore weights r_i, $i = 1,\dots,n$, then (9.15) reduces to (9.14), i.e. $H(\mathbf{p};\mathbf{e}) = H(\mathbf{p})$ for $\mathbf{e} = (1,\dots,1) \in \mathbb{R}^n$.

Remark 9.2 It's well known, when $\mathbf{p} = (p_1,\dots,p_n)$ is a positive probability distribution for some discrete random variable X, i.e. $p_i > 0$, $i = 1,\dots,n$, with $\sum_{i=1}^{n} p_i = 1$, the weighted entropy satisfied estimate

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

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

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

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

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

Shannon's entropy quantifies the unevenness in the probability distribution $\mathbf{p}$.

Example 9.2 For the convex function $f(t) = t\ln t$, $t > 0$, we have

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

and

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i \frac{q_i}{p_i} \ln\left(\frac{q_i}{p_i}\right) = \sum_{i=1}^{n} r_i q_i \ln\left(\frac{q_i}{p_i}\right) = D(\mathbf{q},\mathbf{p};\mathbf{r}). \tag{9.16}$$

We get the Kullback-Leibler divergence or relative entropy as a slight modification of the previous formula for Shannon's entropy (see [19]).
Note that $D(\mathbf{q},\mathbf{p};\mathbf{e}) = D(\mathbf{q},\mathbf{p})$ for $\mathbf{e} = (1,\ldots,1) \in \mathbb{R}^n$.

Remark 9.3 Specially, when $\mathbf{q}$ and $\mathbf{p}$ are two positive probability distributions over the same variable, the Kullback-Leibler divergence is a measure of the difference between them. In statistics, it arises as the expected logarithm of difference between the probability $\mathbf{q}$ of data in the original distribution with the approximating distribution $\mathbf{p}$. It satisfies the following estimates

$$D(\mathbf{q},\mathbf{p}) \geq 0$$

with equality iff $\mathbf{q} = \mathbf{p}$.

Example 9.3 Consider now *the Hellinger distance*

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

where $\mathbf{p},\mathbf{q} \in (0,\infty)^n$. 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 $\mathbf{r} = (r_1,\ldots,r_n) \in [0,\infty)$, 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 $f(t) = \frac{1}{2}\left(1 - \sqrt{t}\right)^2$.

Example 9.4 For the convex function $f(t) = -\sqrt{t}$ and $\mathbf{p},\mathbf{q} \in (0,\infty)^n$, we get

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

known as the Bhattacharyya distance.

Example 9.5 For suitable choices of a convex function f we define:
χ^2-divergence, for $f(t) = (1-t)^2, t > 0$, we have

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

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

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

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

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

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

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

9.3 The Zipf-Mandelbrot entropy

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

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

where

$$H_{n,q,s} = \sum_{i=1}^{n} \frac{1}{(i+q)^s}. \tag{9.17}$$

It is also known as the Pareto-Zipf law, a power-law distribution on ranked data, defined by Mandelbrot [13] as generalization of a simpler distribution called Zipf's law [36]. Many naturally phenomena, as earthquake magnitudes, city sizes, incomes, word frequencies and etc., are distributed according to this distribution. It implies that small occurrences are extremely common, whereas large instances are extremely rare. The Zipf-Mandelbrot has wide applications in many branches of science, as well as linguistics, information sciences , ecological field studies and etc (see for example [8], [25], [26], [35]).

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

$$Z(H,q,s) = \frac{s}{H_{n,q,s}} \sum_{k=1}^{n} \frac{\ln(k+q)}{(k+q)^s} + \ln H_{n,q,s}. \tag{9.18}$$

We also consider the weighted Zipf-Mandelbrot entropy defined by

$$Z(H,q,s,\mathbf{R}) = \frac{s}{H_{n,q,s,\mathbf{R}}} \sum_{k=1}^{n} R_k \frac{\ln(k+q)}{(k+q)^s} + \ln H_{n,q,s,\mathbf{R}} \tag{9.19}$$

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

$$H_{n,q,s,\mathbf{R}} = \sum_{i=1}^{n} \frac{R_i}{(i+q)^s}. \tag{9.20}$$

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

$$H_{n,q,s,\mathbf{r}_j} = \sum_{i=1}^{n} \frac{r_{ij}}{(i+q)^s}. \tag{9.21}$$

9.4 Converse to Sherman's Inequality

We start this section with results including a converse to Sherman's inequality (9.4).

Theorem 9.2 *Let $f : [\alpha,\beta] \to \mathbb{R}$ be a convex function on $[\alpha,\beta] \subset \mathbb{R}$. Let $\mathbf{x} = (x_1,\dots,x_n) \in [\alpha,\beta]^n$, $\mathbf{y} = (y_1,\dots,y_m) \in [\alpha,\beta]^m$, $\mathbf{a} = (a_1,\dots,a_n) \in [0,\infty)^n$ and $\mathbf{b} = (b_1,\dots,b_m) \in [0,\infty)^m$ be such that* (9.3) *holds for some column stochastic matrix $S = (s_{ij}) \in \mathscr{M}_{nm}(\mathbb{R})$. Then*

$$\sum_{j=1}^{m} b_j f(y_j) \le \sum_{i=1}^{n} a_i f(x_i) \le \sum_{j=1}^{m} b_j \frac{f(\alpha)(\beta - y_j) + f(\beta)(y_j - \alpha)}{\beta - \alpha}. \tag{9.22}$$

Proof. Under the assumptions, Sherman's inequality (9.4) holds. Further, from (9.9), setting $p_i = s_{ij}$, for $i = 1,\dots,n$, we have

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

$$
\begin{aligned}
&= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} b_j s_{ij} \right) f(x_i) \\
&= \sum_{j=1}^{m} b_j \left(\sum_{i=1}^{n} s_{ij} f(x_i) \right) \\
&\leq \sum_{j=1}^{m} b_j \left(\frac{\beta - \sum\limits_{i=1}^{n} x_i s_{ij}}{\beta - \alpha} f(\alpha) + \frac{\sum\limits_{i=1}^{n} x_i s_{ij} - \alpha}{\beta - \alpha} f(\beta) \right),
\end{aligned}
$$

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.

Applying Theorem 9.2 we compare two generalized Csiszár f-divergences.

Theorem 9.3 *Let $f : [\alpha, \beta] \to \mathbb{R}$ be a convex function on $[\alpha, \beta] \subset (0, \infty)$. Let $\mathbf{p} \in (0, \infty)^n$, $\mathbf{q} \in (0, \infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha, \beta]$, $i = 1, \ldots, n$. Further, let $\tilde{\mathbf{p}} \in (0, \infty)^m$, $\tilde{\mathbf{q}} \in (0, \infty)^m$, $\mathbf{c} \in [0, \infty)^n$ and $\mathbf{d} \in [0, \infty)^m$ be such that*

$$
\tilde{\mathbf{p}} = \mathbf{p}R, \quad \tilde{\mathbf{q}} = \mathbf{q}R \quad \textit{and} \quad \mathbf{c} = \mathbf{d}R^{\mathsf{T}} \tag{9.23}
$$

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

$$
C_f(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}; \mathbf{d}) \leq C_f(\mathbf{p}, \mathbf{q}; \mathbf{c}) \leq \sum_{j=1}^{m} d_j \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{f(\alpha) \left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} \right) + f(\beta) \left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha \right)}{\beta - \alpha}. \tag{9.24}
$$

Proof. According to (9.11) the inequality (9.24) can be written in the form

$$
\begin{aligned}
\sum_{j=1}^{m} d_j \tilde{p}_j f\left(\frac{\tilde{q}_j}{\tilde{p}_j} \right) &\leq \sum_{i=1}^{n} c_i p_i f\left(\frac{q_i}{p_i} \right) \\
&\leq \sum_{j=1}^{m} d_j \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{f(\alpha) \left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} \right) + f(\beta) \left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha \right)}{\beta - \alpha}.
\end{aligned} \tag{9.25}
$$

We denote $\mathbf{r}_j = (r_{1j}, \ldots, r_{nj})$, $r_{ij} \geq 0$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$. From (9.23) it follows that $\tilde{p}_j = \langle \mathbf{p}, \mathbf{r}_j \rangle = \sum_{i=1}^{n} p_i r_{ij}$ and $\tilde{q}_j = \langle \mathbf{q}, \mathbf{r}_j \rangle = \sum_{i=1}^{n} q_i r_{ij}$ for $j = 1, \ldots, m$. Moreover, $c_i = \sum_{j=1}^{m} d_j r_{ij}$ for $i = 1, \ldots, n$ (see (9.23)) and after multiplying with p_i and taking $a_i = c_i p_i$, $b_j = d_j \langle \mathbf{p}, \mathbf{r}_j \rangle$ we get

$$
a_i = \sum_{j=1}^{m} b_j \frac{p_i r_{ij}}{\langle \mathbf{p}, \mathbf{r}_j \rangle}, \tag{9.26}
$$

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

$$
\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} = \frac{p_1 r_{1j}}{\sum_{i=1}^{n} p_i rij} \frac{q_1}{p_1} + \ldots + \frac{p_n r_{nj}}{\sum_{i=1}^{n} p_i rij} \frac{q_n}{p_n}
$$

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

$$\left[\frac{\langle \mathbf{q},\mathbf{r}_1\rangle}{\langle \mathbf{p},\mathbf{r}_1\rangle},\ldots,\frac{\langle \mathbf{q},\mathbf{r}_m\rangle}{\langle \mathbf{p},\mathbf{r}_m\rangle}\right] = \left[\frac{q_1}{p_1},\ldots,\frac{q_n}{p_n}\right] \begin{bmatrix} \frac{p_1 r_{11}}{\langle \mathbf{p},\mathbf{r}_1\rangle} & \cdots & \frac{p_1 r_{1m}}{\langle \mathbf{p},\mathbf{r}_m\rangle} \\ \vdots & \ddots & \vdots \\ \frac{p_n r_{n1}}{\langle \mathbf{p},\mathbf{r}_1\rangle} & \cdots & \frac{p_n r_{nm}}{\langle \mathbf{p},\mathbf{r}_m\rangle} \end{bmatrix}. \tag{9.27}$$

The $n\times m$ matrix $S = (s_{ij})$, $s_{ij} = \frac{p_i r_{ij}}{\langle \mathbf{p},\mathbf{r}_j\rangle}$ is column stochastic and with $\mathbf{x} = (x_1,\ldots,x_n)$, $\mathbf{y} = (y_1,\ldots,y_n)$, $x_i = \frac{q_i}{p_i}$ and $y_j = \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}$, $i = 1,\ldots,n$, $j = 1,\ldots,m$, satisfies condition $\mathbf{y} = \mathbf{x}S$ (see (11.3.1)). Also, for $\mathbf{a} = (a_1,\ldots,a_n)$ and $\mathbf{b} = (b_1,\ldots,b_m)$, $\mathbf{a} = \mathbf{b}S^\mathsf{T}$ (see (9.26)) is satisfied, so we can apply Theorem 9.2 and obtain

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

which is equivalent to (9.24). □

Corollary 9.1 *Let $f : [\alpha,\beta] \to \mathbb{R}$ be a convex function on $[\alpha,\beta] \subset (0,\infty)$. Let $\mathbf{p} \in (0,\infty)^n$, $\mathbf{q} \in (0,\infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1,\ldots,n$. Further, let $\tilde{\mathbf{p}} \in (0,\infty)^m$ and $\tilde{\mathbf{q}} \in (0,\infty)^m$ be such that*

$$\tilde{\mathbf{p}} = \mathbf{p}R \quad \text{and} \quad \tilde{\mathbf{q}} = \mathbf{q}R \tag{9.28}$$

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

$$C_f(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \le C_f(\mathbf{p},\mathbf{q};\mathbf{R}) \le \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{f(\alpha)\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + f(\beta)\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta-\alpha}. \tag{9.29}$$

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

$$C_f(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \le C_f(\mathbf{p},\mathbf{q}) \le \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{f(\alpha)\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + f(\beta)\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta-\alpha}. \tag{9.30}$$

Proof. By taking $\mathbf{d} = (d_1,\ldots,d_m) = (1,\ldots,1)$ in Theorem 9.3, we calculate $c_i = \sum_{j=1}^m r_{ij} = R_i$ for $i = 1,\ldots,n$. Therefore inequality (9.24) becomes (9.29).
If additionally the matrix R is row stohastic, then $\mathbf{R} = (1,\ldots,1) \in \mathbb{R}^n$ and (9.29) reduces to (9.30). □

As a special case of the previous result we obtain a converse to the Csiszár-Körner inequality (9.12).

Corollary 9.2 *Let $f:[\alpha,\beta]\to\mathbb{R}$ be a convex function on $[\alpha,\beta]\subset(0,\infty)$. Let $\mathbf{p}\in(0,\infty)^n$, $\mathbf{q}\in(0,\infty)^n$, $\mathbf{r}\in[0,\infty)^n$ be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\ldots,n$, with $P_n=\sum_{i=1}^n p_i$ and $Q_n=\sum_{i=1}^n q_i$. Then*

$$\langle\mathbf{p},\mathbf{r}\rangle f\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right)\leq C_f(\mathbf{p},\mathbf{q};\mathbf{r})\leq\langle\mathbf{p},\mathbf{r}\rangle\frac{f(\alpha)\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right)+f(\beta)\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}-\alpha\right)}{\beta-\alpha}. \tag{9.31}$$

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

$$\sum_{i=1}^n p_i f\left(\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i}\right)\leq C_f(\mathbf{p},\mathbf{q})\leq\sum_{i=1}^n p_i\frac{f(\alpha)\left(\beta-\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i}\right)+f(\beta)\left(\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i}-\alpha\right)}{\beta-\alpha}. \tag{9.32}$$

Proof. Taking $m=1$ in Corollary 9.1 and $\mathbf{r}_1=(r_1,\ldots,r_n)$, we obtain $R_i=r_i$ for $i=1,\ldots,n$, and (9.29) becomes (9.31). Further, for $\mathbf{r}=\mathbf{e}=(1,\ldots,1)$, the inequality (9.31) reduces to (9.32). □

9.5 Converses including some entropies and divergences

Using the concept of f-divergence we derive some inequalities for the well-known divergences.

Corollary 9.3 *Let $[\alpha,\beta]\subset(0,\infty)$. Let $\mathbf{p}\in(0,\infty)^n$, $\mathbf{q}\in(0,\infty)^n$ be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\ldots,n$. Further, let $\tilde{\mathbf{p}}\in(0,\infty)^m$, $\tilde{\mathbf{q}}\in(0,\infty)^m$, $\mathbf{c}\in[0,\infty)^n$ and $\mathbf{d}\in[0,\infty)^m$ be such that (9.23) holds for some matrix $R=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$. Then*

$$D(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d})\leq D(\mathbf{p},\mathbf{q};\mathbf{c})\leq\sum_{j=1}^m d_j\langle\mathbf{p},\mathbf{r}_j\rangle\frac{\ln\left(\frac{1}{\alpha}\right)\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+\ln\left(\frac{1}{\beta}\right)\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{9.33}$$

Proof. If we take in Theorem 9.3 function f to be $f(t)=\ln\left(\frac{1}{t}\right)$, which is convex on $[\alpha,\beta]$, then (9.33) follows from (9.24). □

Corollary 9.4 *Let $[\alpha,\beta]\subset(0,\infty)$. Let $\mathbf{p}\in(0,\infty)^n$, $\mathbf{q}\in(0,\infty)^n$, be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\ldots,n$. Let $\tilde{\mathbf{p}}\in(0,\infty)^m$ and $\tilde{\mathbf{q}}\in(0,\infty)^m$ be such that (9.28) 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

$$D(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) \leq D(\mathbf{p}, \mathbf{q}; \mathbf{R}) \leq \sum_{j=1}^m \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{\ln\left(\frac{1}{\alpha}\right)\left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + \ln\left(\frac{1}{\beta}\right)\left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.34}$$

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

$$D(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) \leq D(\mathbf{p}, \mathbf{q}) \leq \sum_{j=1}^m \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{\ln\left(\frac{1}{\alpha}\right)\left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + \ln\left(\frac{1}{\beta}\right)\left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.35}$$

Proof. By taking $\mathbf{d} = (d_1, \ldots, d_m) = (1, \ldots, 1)$ in Theorem 9.3 we obtain $c_i = \sum_{j=1}^m r_{ij} = R_i$ for $i = 1, \ldots, n$. Therefore inequality (9.24) becomes (9.34).
If additionally R is row stohastic, then $\mathbf{R} = (1, \ldots, 1) \in \mathbb{R}^n$ and (9.29) becomes (9.35). □

Corollary 9.5 *Let $[\alpha, \beta] \subset (0, \infty)$. Let $\mathbf{p} \in [\alpha, \beta]^n$, $\tilde{\mathbf{p}} \in [\alpha, \beta]^m$, $\mathbf{c} \in [0, \infty)^n$ and $\mathbf{d} \in [0, \infty)^m$ 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

$$H(\tilde{\mathbf{p}}; \mathbf{d}) \geq H(\mathbf{p}; \mathbf{c}) \geq \sum_{j=1}^m d_j \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{\ln(\alpha)\left(\beta - \frac{1}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + \ln(\beta)\left(\frac{1}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.36}$$

Proof. We take in Theorem 9.3 a function f to be $f(t) = \ln \frac{1}{t}$ which is convex on $[\alpha, \beta]$ and $\mathbf{q} = \mathbf{e} = (1, \ldots, 1) \in \mathbb{R}^m$. Then, since R is column stochastic,we also have $\tilde{\mathbf{q}} = (\langle \mathbf{q}, \mathbf{r}_1 \rangle, \ldots, \langle \mathbf{q}, \mathbf{r}_m \rangle) = (\langle \mathbf{e}, \mathbf{r}_1 \rangle, \ldots, \langle \mathbf{e}, \mathbf{r}_m \rangle) = (1, \ldots, 1)$.Then (9.36) follows from (9.24). □

Corollary 9.6 *Let $[\alpha, \beta] \subset (0, \infty)$. Let $\mathbf{p} \in [\alpha, \beta]^n$ and $\tilde{\mathbf{p}} \in [\alpha, \beta]^m$ be such that*

$$\tilde{\mathbf{p}} = \mathbf{p}R \tag{9.37}$$

for some column stochastic 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

$$H(\tilde{\mathbf{p}}) \geq H(\mathbf{p}; \mathbf{R}) \geq \sum_{j=1}^m \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{\ln(\alpha)\left(\beta - \frac{1}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + \ln(\beta)\left(\frac{1}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.38}$$

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

$$H(\tilde{\mathbf{p}}) \geq H(\mathbf{p}) \geq \sum_{j=1}^m \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{\ln(\alpha)\left(\beta - \frac{1}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + \ln(\beta)\left(\frac{1}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.39}$$

Proof. By taking $\mathbf{d} = (d_1, \dots, d_m) = (1, \dots, 1)$ in Theorem 9.3 we obtain $c_i = \sum_{j=1}^m r_{ij} = R_i$ for $i = 1, \dots, n$. Therefore inequality (9.24) becomes (9.38).
If additionally the matrix R is row stohastic, then $\mathbf{R} = (1, \dots, 1) \in \mathbb{R}^n$ and (9.29) becomes (9.39). □

Corollary 9.7 *Let $[\alpha, \beta] \subset (0, \infty)$. Let $\mathbf{p} \in (0, \infty)^n$, $\mathbf{q} \in (0, \infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha, \beta]$, $i = 1, \dots, n$. Further, let $\tilde{\mathbf{p}} \in (0, \infty)^m$, $\tilde{\mathbf{q}} \in (0, \infty)^m$, $\mathbf{c} \in [0, \infty)^n$ and $\mathbf{d} \in [0, \infty)^m$ be such that (9.23) holds for some matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$. Then*

$$\begin{aligned} h^2(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}; \mathbf{d}) &\leq h^2(\mathbf{p}, \mathbf{q}; \mathbf{c}) \\ &\leq \sum_{j=1}^m d_j \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{(1-\sqrt{\alpha})^2 \left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + (1-\sqrt{\beta})^2 \left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{2(\beta - \alpha)}. \end{aligned} \tag{9.40}$$

Proof. If we take in Theorem 9.3 function f to be $f(t) = \frac{1}{2}\left(1 - \sqrt{t}\right)^2$ which is convex on $[\alpha, \beta]$, equation (9.40) follows from (9.24). □

Corollary 9.8 *Let $[\alpha, \beta] \subset (0, \infty)$. Let $\mathbf{p} \in (0, \infty)^n$, $\mathbf{q} \in (0, \infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha, \beta]$, $i = 1, \dots, n$. Let $\tilde{\mathbf{p}} \in (0, \infty)^m$ and $\tilde{\mathbf{q}} \in (0, \infty)^m$ be such that (9.28) 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*

$$h^2(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) \leq h^2(\mathbf{p}, \mathbf{q}; \mathbf{R}) \leq \sum_{j=1}^m \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{(1-\sqrt{\alpha})^2 \left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + (1-\sqrt{\beta})^2 \left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{2(\beta - \alpha)}. \tag{9.41}$$

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

$$h^2(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}) \leq h^2(\mathbf{p}, \mathbf{q}) \leq \sum_{j=1}^m \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{(1-\sqrt{\alpha})^2 \left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + (1-\sqrt{\beta})^2 \left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{2(\beta - \alpha)}. \tag{9.42}$$

Proof. By taking $\mathbf{d} = (d_1, \dots, d_m) = (1, \dots, 1)$ in Theorem 9.3 we obtain $c_i = \sum_{j=1}^m r_{ij} = R_i$ for $i = 1, \dots, n$. Therefore inequality (9.24) becomes (9.34).
If additionally the matrix R is row stohastic, then $\mathbf{R} = (1, \dots, 1) \in \mathbb{R}^n$ and (9.29) becomes (9.35). □

Corollary 9.9 *Let $[\alpha, \beta] \subset (0, \infty)$. Let $\mathbf{p} \in (0, \infty)^n$, $\mathbf{q} \in (0, \infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha, \beta]$, $i = 1, \dots, n$. Further, let $\tilde{\mathbf{p}} \in (0, \infty)^m$, $\tilde{\mathbf{q}} \in (0, \infty)^m$, $\mathbf{c} \in [0, \infty)^n$ and $\mathbf{d} \in [0, \infty)^m$ be such that (9.23) holds for some matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$. Then*

$$B(\tilde{\mathbf{p}}, \tilde{\mathbf{q}}; \mathbf{d}) \geq B(\mathbf{p}, \mathbf{q}; \mathbf{c}) \geq \sum_{j=1}^m d_j \langle \mathbf{p}, \mathbf{r}_j \rangle \frac{\sqrt{\alpha} \left(\beta - \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) + \sqrt{\beta} \left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.43}$$

Proof. If we take in Theorem 9.3 function f to be $f(t) = -\sqrt{t}$, which is convex on $[\alpha,\beta]$, equation (9.43) follows from (9.24). □

Corollary 9.10 *Let $[\alpha,\beta] \subset (0,\infty)$. Let $\mathbf{p} \in (0,\infty)^n$, $\mathbf{q} \in (0,\infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1,\ldots,n$. Let $\tilde{\mathbf{p}} \in (0,\infty)^m$ and $\tilde{\mathbf{q}} \in (0,\infty)^m$ be such that (9.28) 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*

$$B(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \geq B(\mathbf{p},\mathbf{q};\mathbf{R}) \geq \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\sqrt{\alpha}\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + \sqrt{\beta}\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.44}$$

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

$$B(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \geq B(\mathbf{p},\mathbf{q}) \geq \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\sqrt{\alpha}\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + \sqrt{\beta}\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.45}$$

Proof. By taking $\mathbf{d} = (d_1,\ldots,d_m) = (1,\ldots,1)$ in Theorem 9.3 we obtain $c_i = \sum_{j=1}^m r_{ij} = R_i$ for $i = 1,\ldots,n$. Therefore inequality (9.24) becomes (9.44).
If additionally R is row stohastic, then $\mathbf{R} = (1,\ldots,1) \in \mathbb{R}^n$ and (9.29) becomes (9.45). □

Corollary 9.11 *Let $[\alpha,\beta] \subset (0,\infty)$. Let $\mathbf{p} \in (0,\infty)^n$, $\mathbf{q} \in (0,\infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1,\ldots,n$. Further, let $\tilde{\mathbf{p}} \in (0,\infty)^m$, $\tilde{\mathbf{q}} \in (0,\infty)^m$, $\mathbf{c} \in [0,\infty)^n$ and $\mathbf{d} \in [0,\infty)^m$ be such that (9.23) holds for some matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$. Then*

$$\begin{aligned}\chi^2(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d}) &\leq \chi^2(\mathbf{p},\mathbf{q};\mathbf{c}) \\ &\leq \sum_{j=1}^m d_j \langle \mathbf{p},\mathbf{r}_j\rangle \frac{(1-\alpha)^2\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + (1-\beta)^2\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}.\end{aligned} \tag{9.46}$$

Proof. If we take in Theorem 9.3 function f to be $f(t) = (1-t)^2$ which is convex on $[\alpha,\beta]$, equation (9.46) follows from (9.24). □

Corollary 9.12 *Let $[\alpha,\beta] \subset (0,\infty)$. Let $\mathbf{p} \in (0,\infty)^n$, $\mathbf{q} \in (0,\infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1,\ldots,n$. Let $\tilde{\mathbf{p}} \in (0,\infty)^m$ and $\tilde{\mathbf{q}} \in (0,\infty)^m$ be such that (9.28) 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*

$$\chi^2(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leq \chi^2(\mathbf{p},\mathbf{q};\mathbf{R}) \leq \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{(1-\alpha)^2\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + (1-\beta)^2\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.47}$$

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

$$\chi^2(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \leq \chi^2(\mathbf{p},\mathbf{q}) \leq \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{(1-\alpha)^2\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + (1-\beta)^2\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta - \alpha}. \tag{9.48}$$

Proof. By taking $\mathbf{d} = (d_1,\ldots,d_m) = (1,\ldots,1)$ in Theorem 9.3 we obtain $c_i = \sum_{j=1}^m r_{ij} = R_i$ for $i = 1,\ldots,n$. Therefore inequality (9.24) becomes (9.47).
If additionally R is row stohastic, then $\mathbf{R} = (1,\ldots,1) \in \mathbb{R}^n$ and (9.29) becomes (9.48). □

Corollary 9.13 *Let $[\alpha,\beta] \subset (0,\infty)$. Let $\mathbf{p} \in (0,\infty)^n$, $\mathbf{q} \in (0,\infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1,\ldots,n$. Further, let $\tilde{\mathbf{p}} \in (0,\infty)^m$, $\tilde{\mathbf{q}} \in (0,\infty)^m$, $\mathbf{c} \in [0,\infty)^n$ and $\mathbf{d} \in [0,\infty)^m$ be such that (9.23) holds for some matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$. Then*

$$\begin{aligned} V(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d}) &\le V(\mathbf{p},\mathbf{q};\mathbf{c}) \\ &\le \sum_{j=1}^m d_j \langle \mathbf{p},\mathbf{r}_j\rangle \frac{|1-\alpha|\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + |1-\beta|\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta-\alpha}. \end{aligned} \tag{9.49}$$

Proof. If we take in Theorem 9.3 function f to be $f(t) = |1-t|$ which is convex on $[\alpha,\beta]$, equation (9.49) follows from (9.24). □

Corollary 9.14 *Let $[\alpha,\beta] \subset (0,\infty)$. Let $\mathbf{p} \in (0,\infty)^n$, $\mathbf{q} \in (0,\infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1,\ldots,n$. Let $\tilde{\mathbf{p}} \in (0,\infty)^m$ and $\tilde{\mathbf{q}} \in (0,\infty)^m$ be such that (9.28) 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*

$$V(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \le V(\mathbf{p},\mathbf{q};\mathbf{R}) \le \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{|1-\alpha|\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + |1-\beta|\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta-\alpha}. \tag{9.50}$$

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

$$V(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \le V(\mathbf{p},\mathbf{q}) \le \sum_{j=1}^m \langle \mathbf{p},\mathbf{r}_j\rangle \frac{|1-\alpha|\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + |1-\beta|\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta-\alpha}. \tag{9.51}$$

Proof. By taking $\mathbf{d} = (d_1,\ldots,d_m) = (1,\ldots,1)$ in Theorem 9.3 we obtain $c_i = \sum_{j=1}^m r_{ij} = R_i$ for $i = 1,\ldots,n$. Therefore inequality (9.24) becomes (9.50).
If additionally R is row stohastic, then $\mathbf{R} = (1,\ldots,1) \in \mathbb{R}^n$ and (9.29) becomes (9.51). □

Corollary 9.15 *Let $[\alpha,\beta] \subset (0,\infty)$. Let $\mathbf{p} \in (0,\infty)^n$, $\mathbf{q} \in (0,\infty)^n$, be such that $\frac{q_i}{p_i} \in [\alpha,\beta]$, $i = 1,\ldots,n$. Further, let $\tilde{\mathbf{p}} \in (0,\infty)^m$, $\tilde{\mathbf{q}} \in (0,\infty)^m$, $\mathbf{c} \in [0,\infty)^n$ and $\mathbf{d} \in [0,\infty)^m$ be such that (9.23) holds for some matrix $R = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$. Then*

$$\Delta(\tilde{\mathbf{p}},\tilde{\mathbf{q}};\mathbf{d}) \le \Delta(\mathbf{p},\mathbf{q};\mathbf{c}) \le \sum_{j=1}^m d_j \langle \mathbf{p},\mathbf{r}_j\rangle \frac{\frac{(1-\alpha)^2}{\alpha+1}\left(\beta - \frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle}\right) + \frac{(1-\beta)^2}{\beta+1}\left(\frac{\langle \mathbf{q},\mathbf{r}_j\rangle}{\langle \mathbf{p},\mathbf{r}_j\rangle} - \alpha\right)}{\beta-\alpha}. \tag{9.52}$$

Proof. If we take in Theorem 9.3 function f to be $f(t) = \frac{(1-t)^2}{t+1}$ which is convex on $[\alpha,\beta]$, equation (9.52) follows from (9.24). □

Corollary 9.16 *Let $[\alpha,\beta]\subset(0,\infty)$. Let $\mathbf{p}\in(0,\infty)^n$, $\mathbf{q}\in(0,\infty)^n$, be such that $\frac{q_i}{p_i}\in[\alpha,\beta]$, $i=1,\ldots,n$. Let $\tilde{\mathbf{p}}\in(0,\infty)^m$ and $\tilde{\mathbf{q}}\in(0,\infty)^m$ be such that (9.28) 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*

$$\Delta(\tilde{\mathbf{p}},\tilde{\mathbf{q}})\leq\Delta(\mathbf{p},\mathbf{q};\mathbf{R})\leq\sum_{j=1}^m\langle\mathbf{p},\mathbf{r}_j\rangle\frac{\frac{(1-\alpha)^2}{\alpha+1}\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+\frac{(1-\beta)^2}{\beta+1}\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{9.53}$$

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

$$\Delta(\tilde{\mathbf{p}},\tilde{\mathbf{q}})\leq\Delta(\mathbf{p},\mathbf{q})\leq\sum_{j=1}^m\langle\mathbf{p},\mathbf{r}_j\rangle\frac{\frac{(1-\alpha)^2}{\alpha+1}\left(\beta-\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)+\frac{(1-\beta)^2}{\beta+1}\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}-\alpha\right)}{\beta-\alpha}. \tag{9.54}$$

Proof. By taking $\mathbf{d}=(d_1,\ldots,d_m)=(1,\ldots,1)$ in Theorem 9.3 we obtain $c_i=\sum_{j=1}^m r_{ij}=R_i$ for $i=1,\ldots,n$. Therefore inequality (9.24) becomes (9.53).
If additionally R is row stohastic, then $\mathbf{R}=(1,\ldots,1)\in\mathbb{R}^n$ and (9.29) becomes (9.54). □

9.6 Converses including Zipf-Mandelbrot entropy

Here we give some inequalities including the Zipf-Mandelbrot entropies (9.18) and (9.19).

Theorem 9.4 *Let $n\in\mathbb{N}$, $q\geq0$ 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*

$$\sum_{j=1}^m\frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s,\mathbf{R}}}\ln\left(\frac{H_{n,q,s,\mathbf{R}}}{H_{n,q,s,\mathbf{r}_j}}\right)\geq Z(H,q,s,\mathbf{R}) \tag{9.55}$$

$$\geq\sum_{j=1}^m\frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s,\mathbf{R}}}\frac{\ln(\alpha)\left(\beta-\frac{H_{n,q,s,\mathbf{R}}}{H_{n,q,s,\mathbf{r}_j}}\right)+\ln(\beta)\left(\frac{H_{n,q,s,\mathbf{R}}}{H_{n,q,s,\mathbf{r}_j}}-\alpha\right)}{\beta-\alpha},$$

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

$$\sum_{j=1}^m\frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s}}\ln\left(\frac{H_{n,q,s}}{H_{n,q,s,\mathbf{r}_j}}\right)\geq Z(H,q,s) \tag{9.56}$$

$$\geq\sum_{j=1}^m\frac{H_{n,q,s,\mathbf{r}_j}}{H_{n,q,s}}\frac{\ln(\alpha)\left(\beta-\frac{H_{n,q,s}}{H_{n,q,s,\mathbf{r}_j}}\right)+\ln(\beta)\left(\frac{H_{n,q,s}}{H_{n,q,s,\mathbf{r}_j}}-\alpha\right)}{\beta-\alpha}.$$

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 $\frac{1}{(i+q)^s H_{n,q,s,\mathbf{R}}}$, $i = 1,2,\dots,n$, 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 (9.38) we get the required result.
Specially, if R is also row stochastic, then $\mathbf{R} = (1,\dots,1) \in \mathbb{R}^n$. 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 (9.55) reduces to (9.56). □

9.7 Bounds for Sherman's difference

Considering the difference between the right and left side of Sherman's inequality (9.4) we may Sherman's inequality write in the form

$$0 \leq \sum_{i=1}^{n} a_i f(x_i) - \sum_{j=1}^{m} b_j f(y_j). \tag{9.57}$$

Here we present new lower and upper bounds for Sherman's difference $\sum_{i=1}^{n} a_i f(x_i) - \sum_{j=1}^{m} b_j f(y_j)$.

Theorem 9.5 *Let $[\alpha,\beta] \subset \mathbb{R}$, $\mathbf{x} = (x_1,\ldots,x_n) \in [\alpha,\beta]^n$, $\mathbf{y} = (y_1,\ldots,y_m) \in [\alpha,\beta]^m$, $\mathbf{a} = (a_1,\ldots,a_n) \in [0,\infty)^n$ and $\mathbf{b} = (b_1,\ldots,b_m) \in [0,\infty)^m$ be such that*

$$\mathbf{y} = \mathbf{x}S \quad \textit{and} \quad \mathbf{a} = \mathbf{b}S^T \tag{9.58}$$

holds for some column stochastic matrix $S = (s_{ij}) \in \mathscr{M}_{nm}(\mathbb{R})$. Then for every convex function $f : [\alpha,\beta] \to \mathbb{R}$, the inequality

$$0 \leqslant \min_{1\leqslant i\leqslant n}\{a_i\} S_f(\mathbf{x}) \leqslant \sum_{i=1}^{n} a_i f(x_i) - \sum_{j=1}^{m} b_j f(y_j) \leqslant \max_{1\leqslant i\leqslant n}\{a_i\} S_f(\mathbf{x}) \tag{9.59}$$

holds, where $S_f(\mathbf{x})$ is defined by (9.8).

Proof. Under assumptions (9.58), i.e. $y_j = \sum_{i=1}^{n} x_i s_{ij}$, $j = 1,\ldots,m$ and $a_i = \sum_{j=1}^{m} b_j s_{ij}$, $i = 1,\ldots,n$, we have

$$\begin{aligned}\sum_{i=1}^{n} a_i f(x_i) - \sum_{j=1}^{m} b_j f(y_j) &= \sum_{j=1}^{m} b_j \sum_{i=1}^{n} s_{ij} f(x_i) - \sum_{j=1}^{m} b_j f\left(\sum_{i=1}^{n} x_i s_{ij}\right)\\ &= \sum_{j=1}^{m} b_j \left(\sum_{i=1}^{n} s_{ij} f(x_i) - f\left(\sum_{i=1}^{n} x_i s_{ij}\right)\right).\end{aligned} \tag{9.60}$$

Applying Theorem 1 to (9.60) we get

$$0 \leqslant \sum_{j=1}^{m} b_j \min_{1\leqslant i\leqslant n}\{s_{ij}\} S_f(\mathbf{x}) \leqslant \sum_{i=1}^{n} a_i f(x_i) - \sum_{j=1}^{m} b_j f(y_j) \leqslant \sum_{j=1}^{m} b_j \max_{1\leqslant i\leqslant n}\{s_{ij}\} S_f(\mathbf{x}),$$

where $S_f(\mathbf{x})$ is defined by (9.8). Moreover, we have

$$0 \leqslant \min_{1\leqslant i\leqslant n}\{a_i\} S_f(\mathbf{x}) \leqslant \sum_{i=1}^{n} a_i f(x_i) - \sum_{j=1}^{m} b_j f(y_j) \leqslant \max_{1\leqslant i\leqslant n}\{a_i\} S_f(\mathbf{x}),$$

what we need to prove. □

Remark 9.4 The previous theorem presents generalizations of Theorem 9.1. Choosing $m = 1$ and setting $\mathbf{b} = [1]$, the inequality (9.59) reduces to the form (9.7).

Theorem 9.6 *Let $f : [0,\infty) \to \mathbb{R}$ be a convex function on $[0,\infty)$. Let $\mathbf{p} = (p_1,\dots,p_n) \in (0,\infty)^n$, $\mathbf{q} = (q_1,\dots,q_n) \in (0,\infty)^n$, $\mathbf{b} = (b_1,\dots,b_m) \in [0,\infty)^m$ and $\mathbf{R} = (r_{ij}) \in \mathscr{M}_{nm}(\mathbb{R}_+)$. Let us define*

$$\langle \mathbf{p}, \mathbf{r}_j \rangle = \sum_{i=1}^{n} p_i r_{ij} > 0, \quad \langle \mathbf{q}, \mathbf{r}_j \rangle = \sum_{i=1}^{n} q_i r_{ij}, \quad j = 1,\dots,m,$$

$$a_i = \sum_{j=1}^{m} b_j \frac{p_i r_{ij}}{\langle \mathbf{p}, \mathbf{r}_j \rangle}, \quad i = 1,\dots,n,$$

$$S_f(\mathbf{q},\mathbf{p}) = \sum_{i=1}^{n} f\left(\frac{q_i}{p_i}\right) - nf\left(\frac{1}{n}\sum_{i=1}^{n}\frac{q_i}{p_i}\right) \geq 0. \tag{9.61}$$

Then

$$0 \leqslant \min_{1\leqslant i\leqslant n}\{a_i\} S_f(\mathbf{q},\mathbf{p}) \leqslant \sum_{i=1}^{n} a_i f\left(\frac{q_i}{p_i}\right) - \sum_{j=1}^{m} b_j f\left(\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}\right) \leqslant \max_{1\leqslant i\leqslant n}\{a_i\} S_f(\mathbf{q},\mathbf{p}). \tag{9.62}$$

Proof. Let us consider $\mathbf{x} = (x_1,\dots,x_n)$ and $\mathbf{y} = (y_1,\dots,y_m)$, such that $x_i = \frac{q_i}{p_i}$, $i = 1,\dots,n$ and $y_j = \frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle}$, $j = 1,\dots,m$.
The following equality holds:

$$\frac{\langle \mathbf{q}, \mathbf{r}_j \rangle}{\langle \mathbf{p}, \mathbf{r}_j \rangle} = \frac{\sum_{i=1}^{n} q_i r_{ij}}{\sum_{i=1}^{n} p_i r_{ij}} = \frac{p_1 r_{1j}}{\sum_{i=1}^{n} p_i r_{ij}}\frac{q_1}{p_1} + \dots + \frac{p_n r_{nj}}{\sum_{i=1}^{n} p_i r_{ij}}\frac{q_n}{p_n}, \quad j = 1,\dots,m.$$

Moreover, the following identity

$$\left(\frac{\langle \mathbf{q}, \mathbf{r}_1 \rangle}{\langle \mathbf{p}, \mathbf{r}_1 \rangle},\dots,\frac{\langle \mathbf{q}, \mathbf{r}_m \rangle}{\langle \mathbf{p}, \mathbf{r}_m \rangle}\right) = \left(\frac{q_1}{p_1},\dots,\frac{q_n}{p_n}\right) \cdot \begin{pmatrix} \frac{p_1 r_{11}}{\langle \mathbf{p}, \mathbf{r}_1 \rangle} & \cdots & \frac{p_1 r_{1m}}{\langle \mathbf{p}, \mathbf{r}_m \rangle} \\ \vdots & \ddots & \vdots \\ \frac{p_n r_{n1}}{\langle \mathbf{p}, \mathbf{r}_1 \rangle} & \cdots & \frac{p_n r_{nm}}{\langle \mathbf{p}, \mathbf{r}_m \rangle} \end{pmatrix}$$

is valid for some column stochastic matrix $S = (s_{ij}) \in \mathscr{M}_{nm}(\mathbb{R})$, with $s_{ij} = \frac{p_i r_{ij}}{\langle \mathbf{p}, \mathbf{r}_j \rangle}$, $i = 1,\dots,n$, $j = 1,\dots,m$. Therefore, $\mathbf{y} = \mathbf{x}S$ holds.
Further, by definition $a_i = \sum_{j=1}^{m} b_j \frac{p_i r_{ij}}{\langle \mathbf{p}, \mathbf{r}_j \rangle}$, $i = 1,\dots,n$, i.e. $\mathbf{a} = \mathbf{b}S^T$. Therefore, the assumptions of Theorem 9.5 are fulfill. Now applying (9.59) we get the required result. □

9.8 Bounds for Csiszár f-divergence functional

The followig results give new bounds for f-divergence functionals (9.10) and (9.11).

Corollary 9.17 *Let $f:[0,\infty)\to\mathbb{R}$ be a convex function on $[0,\infty)$. Let $\mathbf{p}=(p_1,\ldots,p_n)\in(0,\infty)^n$, $\mathbf{q}=(q_1,\ldots,q_n)\in(0,\infty)^n$, $\mathbf{R}=(r_{ij})\in\mathscr{M}_{nm}(\mathbb{R}_+)$ and $\tilde{\mathbf{R}}=(R_1,\ldots,R_n)$, with $R_i=\sum_{j=1}^{m} r_{ij}$. Let $C_f(\mathbf{q},\mathbf{p})$, $C_f(\mathbf{q},\mathbf{p};\tilde{\mathbf{R}})$, $\langle\mathbf{p},\mathbf{r}_j\rangle$, $\langle\mathbf{q},\mathbf{r}_j\rangle$ and $S_f(\mathbf{q},\mathbf{p})$ be defined as in (9.10), (9.11) and (9.61), respectively. Then*

$$\begin{aligned}0 &\leqslant \min_{1\leqslant i\leqslant n}\{p_iR_i\}S_f(\mathbf{q},\mathbf{p}) \\ &\leqslant C_f(\mathbf{q},\mathbf{p};\tilde{\mathbf{R}})-\sum_{j=1}^{m}\langle\mathbf{p},\mathbf{r}_j\rangle f\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)\leqslant \max_{1\leqslant i\leqslant n}\{p_iR_i\}S_f(\mathbf{q},\mathbf{p}).\end{aligned} \tag{9.63}$$

If in addition $\mathbf{R}$ is row stochastic, then

$$\begin{aligned}0 &\leqslant \min_{1\leqslant i\leqslant n}\{p_i\}S_f(\mathbf{q},\mathbf{p}) \\ &\leqslant C_f(\mathbf{q},\mathbf{p})-\sum_{j=1}^{m}\langle\mathbf{p},\mathbf{r}_j\rangle f\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right)\leqslant \max_{1\leqslant i\leqslant n}\{p_i\}S_f(\mathbf{q},\mathbf{p}).\end{aligned} \tag{9.64}$$

Proof. Applying (9.62) with substitution b_j with $\langle\mathbf{p},\mathbf{r}_j\rangle$ for $j=1,\ldots,m$, from which it follows

$$a_i=\sum_{j=1}^{m} b_j\frac{p_ir_{ij}}{\langle\mathbf{p},\mathbf{r}_j\rangle}=p_i\sum_{j=1}^{m} r_{ij}=p_iR_i,\quad i=1,\ldots,n,$$

we get

$$\begin{aligned}0 \leqslant \min_{1\leqslant i\leqslant n}\{p_iR_i\}S_f(\mathbf{q},\mathbf{p}) &\leqslant \sum_{i=1}^{n} p_iR_i f\left(\frac{q_i}{p_i}\right)-\sum_{j=1}^{m}\langle\mathbf{p},\mathbf{r}_j\rangle f\left(\frac{\langle\mathbf{q},\mathbf{r}_j\rangle}{\langle\mathbf{p},\mathbf{r}_j\rangle}\right) \\ &\leqslant \max_{1\leqslant i\leqslant n}\{p_iR_i\}S_f(\mathbf{q},\mathbf{p})\end{aligned} \tag{9.65}$$

which is equivalent to (9.63).
If in addition the matrix $\mathbf{R}$ is row stochastic, i.e. $R_i=\sum_{j=1}^{m} r_{ij}=1$ for all $i=1,\ldots,n$, then (9.65) becomes

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

which is equivalent to (9.64). □

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

Corollary 9.18 *Let $f:[0,\infty)\to\mathbb{R}$ be a convex function on $[0,\infty)$. Let $\mathbf{p}=(p_1,\dots,p_n)\in(0,\infty)^n$, $\mathbf{q}=(q_1,\dots,q_n)\in(0,\infty)^n$ and $\mathbf{r}=(r_1,\dots,r_n)\in[0,\infty)^n$. Let us define*

$$\langle\mathbf{p},\mathbf{r}\rangle=\sum_{i=1}^{n}p_ir_i>0,\quad \langle\mathbf{q},\mathbf{r}\rangle=\sum_{i=1}^{n}q_ir_i.$$

Let $C_f(\mathbf{q},\mathbf{p})$, $C_f(\mathbf{q},\mathbf{p};\mathbf{r})$ and $S_f(\mathbf{q},\mathbf{p})$ be define by (9.10), (9.11) and (9.61), respectively. Then

$$\begin{aligned}0&\leqslant\min_{1\leqslant i\leqslant n}\{p_ir_i\}S_f(\mathbf{q},\mathbf{p})\\&\leqslant C_f(\mathbf{q},\mathbf{p};\mathbf{r})-\langle\mathbf{p},\mathbf{r}\rangle f\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right)\leqslant\max_{1\leqslant i\leqslant n}\{p_ir_i\}S_f(\mathbf{q},\mathbf{p}).\end{aligned}\tag{9.66}$$

If in addition $\mathbf{r}=\mathbf{e}=(1,\dots,1)$, then

$$\begin{aligned}0&\leqslant\min_{1\leqslant i\leqslant n}\{p_i\}S_f(\mathbf{q},\mathbf{p})\\&\leqslant C_f(\mathbf{q},\mathbf{p})-\sum_{i=1}^{n}p_if\left(\frac{\sum_{i=1}^{n}q_i}{\sum_{i=1}^{n}p_i}\right)\leqslant\max_{1\leqslant i\leqslant n}\{p_i\}S_f(\mathbf{q},\mathbf{p}).\end{aligned}\tag{9.67}$$

Remark 9.5 (i) Note that the Csizar-Korner inequality (9.12) is generalized and refined in (9.66) and (9.67). Further, the inequality (9.66) is equivalent to

$$\begin{aligned}\langle\mathbf{p},\mathbf{r}\rangle f\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right)&\leqslant\min_{1\leqslant i\leqslant n}\{p_ir_i\}S_f(\mathbf{q},\mathbf{p})+\langle\mathbf{p},\mathbf{r}\rangle f\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right)\\&\leqslant C_f(\mathbf{q},\mathbf{p};\mathbf{r})\\&\leqslant\max_{1\leqslant i\leqslant n}\{p_ir_i\}S_f(\mathbf{q},\mathbf{p})+\langle\mathbf{p},\mathbf{r}\rangle f\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right)\end{aligned}$$

and the inequality (9.67) to

$$\begin{aligned}P_nf\left(\frac{Q_n}{P_n}\right)&\leqslant\min_{1\leqslant i\leqslant n}\{p_i\}S_f(\mathbf{q},\mathbf{p})+P_nf\left(\frac{Q_n}{P_n}\right)\\&\leqslant C_f(\mathbf{q},\mathbf{p})\\&\leqslant\max_{1\leqslant i\leqslant n}\{p_i\}S_f(\mathbf{q},\mathbf{p})+P_nf\left(\frac{Q_n}{P_n}\right),\end{aligned}$$

where $P_n=\sum_{i=1}^{n}p_i$ and $Q_n=\sum_{i=1}^{n}q_i$. Specially, if f is normalized, i.e. $f(1)=0$ and $\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i$, we get the lower and upper bounds for Csiszár f-divergence in the form

$$0\leqslant\min_{1\leqslant i\leqslant n}\{p_i\}S_f(\mathbf{q},\mathbf{p})\leqslant C_f(\mathbf{q},\mathbf{p})\leqslant\max_{1\leqslant i\leqslant n}\{p_i\}S_f(\mathbf{q},\mathbf{p}).\tag{9.68}$$

(ii) Taking $\mathbf{p} = \mathbf{e} = (1,1,\ldots,1)$, from (9.66) we obtain the bounds for Jensen's functional in the form

$$0 \leq \min_{1\leqslant i\leqslant n}\{r_i\}\left[\sum_{i=1}^{n} f(q_i) - nf\left(\frac{1}{n}\sum_{i=1}^{n} q_i\right)\right] \leqslant \sum_{i=1}^{n} r_i f(q_i) - R_n f\left(\frac{1}{R_n}\sum_{i=1}^{n} q_i r_i\right) \tag{9.69}$$
$$\leqslant \max_{1\leqslant i\leqslant n}\{r_i\}\left[\sum_{i=1}^{n} f(q_i) - nf\left(\frac{1}{n}\sum_{i=1}^{n} q_i\right)\right],$$

where $R_n = \sum\limits_{i=1}^{n} r_i > 0$.

(iii) Further, let us denote $Q_n = \sum_{i=1}^{n} q_i > 0$. If we substitute r_i with $\frac{q_i}{Q_n}$ and q_i with $\frac{p_i}{q_i}$, then from (9.69) we get

$$\min_{1\leqslant i\leqslant n}\{q_i\}\left[\sum_{i=1}^{n} f\left(\frac{p_i}{q_i}\right) - nf\left(\frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{q_i}\right)\right] \leqslant \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) - Q_n R_n f\left(\frac{1}{R_n Q_n}\sum_{i=1}^{n} p_i\right) \tag{9.70}$$
$$\leqslant \max_{1\leqslant i\leqslant n}\{q_i\}\left[\sum_{i=1}^{n} f\left(\frac{p_i}{q_i}\right) - nf\left(\frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{q_i}\right)\right].$$

9.9 Bounds for some entropies and divergences

Applying results from the previous section, we estimate some new bounds for some well known entropies. In the following results we use notation $\langle\cdot,\cdot\rangle$ for the standard inner product.

Theorem 9.7 *Let* $\mathbf{p}$ *be positive probability distributions,* $\mathbf{q} = (q_1,\ldots,q_n) \in (0,\infty)^n$ *and* $\mathbf{r} = (r_1,\ldots,r_n) \in [0,\infty)^n$. *Then*

$$\sum_{i=1}^{n} r_i p_i \ln\frac{1}{q_i} + \langle\mathbf{p},\mathbf{r}\rangle \ln\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right) - \max_{1\leqslant i\leqslant n}\{p_i r_i\}\tilde{S}(\mathbf{q},\mathbf{p}) \tag{9.71}$$
$$\leqslant H(\mathbf{p};\mathbf{r})$$
$$\leqslant \sum_{i=1}^{n} r_i p_i \ln\frac{1}{q_i} + \langle\mathbf{p},\mathbf{r}\rangle \ln\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right) - \min_{1\leqslant i\leqslant n}\{p_i r_i\}\tilde{S}(\mathbf{q},\mathbf{p})$$
$$\leqslant \sum_{i=1}^{n} r_i p_i \ln\frac{1}{q_i} + \langle\mathbf{p},\mathbf{r}\rangle \ln\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right),$$

where

$$\tilde{S}(\mathbf{q},\mathbf{p}) = -\sum_{i=1}^{n} \ln\left(\frac{q_i}{p_i}\right) + n\ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{q_i}{p_i}\right) \geq 0.$$

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

$$\begin{aligned}\sum_{i=1}^{n} p_i \ln\frac{1}{q_i}+\ln\lambda-\max_{1\leqslant i\leqslant n}\{p_i\}\tilde{S}(\mathbf{q},\mathbf{p}) &\leqslant H(\mathbf{p}) \\ &\leqslant \sum_{i=1}^{n} p_i \ln\frac{1}{q_i}+\ln\lambda-\min_{1\leqslant i\leqslant n}\{p_i\}\tilde{S}(\mathbf{q},\mathbf{p}) \\ &\leqslant \sum_{i=1}^{n} p_i \ln\frac{1}{q_i}+\ln\lambda,\end{aligned} \tag{9.72}$$

Proof. Applying (9.66) to the convex function $f(t)=-\ln t$, we get

$$\begin{aligned}0 &\leqslant \min_{1\leqslant i\leqslant n}\{p_i r_i\}\tilde{S}(\mathbf{q},\mathbf{p}) \\ &\leqslant \sum_{i=1}^{n} r_i p_i \ln\frac{1}{q_i}-H(\mathbf{p};\mathbf{r})+\langle\mathbf{p},\mathbf{r}\rangle\ln\left(\frac{\langle\mathbf{q},\mathbf{r}\rangle}{\langle\mathbf{p},\mathbf{r}\rangle}\right)\leqslant \max_{1\leqslant i\leqslant n}\{p_i r_i\}\tilde{S}(\mathbf{q},\mathbf{p}).\end{aligned} \tag{9.73}$$

where $\tilde{S}(\mathbf{q},\mathbf{p})=-\sum_{i=1}^{n}\ln\left(\frac{q_i}{p_i}\right)+n\ln\left(\frac{1}{n}\sum_{i=1}^{n}\frac{q_i}{p_i}\right)\geq 0$, which is equivalent to (9.71). Further, choosing $\mathbf{r}=\mathbf{e}=(1,\ldots,1)$ and setting $\sum_{i=1}^{n} q_i=\lambda$, we have

$$\langle\mathbf{p},\mathbf{r}\rangle=\langle\mathbf{p},\mathbf{e}\rangle=\sum_{i=1}^{n} p_i=1,\quad \langle\mathbf{q},\mathbf{r}\rangle=\langle\mathbf{q},\mathbf{e}\rangle=\sum_{i=1}^{n} q_i=\lambda,$$

i.e. (9.73) reduces to

$$0\leqslant \min_{1\leqslant i\leqslant n}\{p_i\}\tilde{S}(\mathbf{q},\mathbf{p})\leqslant \sum_{i=1}^{n} p_i\ln\frac{1}{q_i}-H(\mathbf{p})-\ln\lambda\leqslant \max_{1\leqslant i\leqslant n}\{p_i\}\tilde{S}(\mathbf{q},\mathbf{p}),$$

which is equivalent to (9.72). □

Corollary 9.19 *Let* $\mathbf{p}$ *be a positive probability distributions and* $\mathbf{r}=(r_1,\ldots,r_n)\in[0,\infty)^n$ *with* $R_n=\sum_{i=1}^{n} r_i$. *Then*

$$\begin{aligned}\langle\mathbf{p},\mathbf{r}\rangle\ln\left(\frac{R_n}{\langle\mathbf{p},\mathbf{r}\rangle}\right)-\max_{1\leqslant i\leqslant n}\{p_i r_i\}\tilde{S}(\mathbf{e},\mathbf{p}) &\leqslant H(\mathbf{p};\mathbf{r}) \\ &\leqslant \langle\mathbf{p},\mathbf{r}\rangle\ln\left(\frac{R_n}{\langle\mathbf{p},\mathbf{r}\rangle}\right)-\min_{1\leqslant i\leqslant n}\{p_i r_i\}\tilde{S}(\mathbf{e},\mathbf{p}) \\ &\leqslant \langle\mathbf{p},\mathbf{r}\rangle\ln\left(\frac{R_n}{\langle\mathbf{p},\mathbf{r}\rangle}\right),\end{aligned} \tag{9.74}$$

where

$$\tilde{S}(\mathbf{e},\mathbf{p})=\sum_{i=1}^{n}\ln p_i-n\ln n\geq 0.$$

If in addition $\mathbf{r} = \mathbf{e} = (1, \ldots, 1)$, *then*

$$\begin{aligned}\ln n - \max_{1 \leqslant i \leqslant n} \{p_i\} \tilde{S}(\mathbf{e}, \mathbf{p}) &\leqslant H(\mathbf{p}) \\ &\leqslant \ln n - \min_{1 \leqslant i \leqslant n} \{p_i\} \tilde{S}(\mathbf{e}, \mathbf{p}) \\ &\leqslant \ln n,\end{aligned} \tag{9.75}$$

Proof. Setting $\mathbf{q} = \mathbf{e} = (1, \ldots, 1)$, the inequality (9.71) reduces to (9.74). If in addition $\mathbf{r} = \mathbf{e} = (1, \ldots, 1)$, then (9.74) becomes (9.75). □

Theorem 9.8 *Let* $\mathbf{p}$ *and* $\mathbf{q}$ *be two positive probability distributions and* $\mathbf{r} = (r_1, \ldots, r_n) \in [0, \infty)^n$. *Then*

$$\begin{aligned}\langle \mathbf{q}, \mathbf{r} \rangle \ln \left(\frac{\langle \mathbf{q}, \mathbf{r} \rangle}{\langle \mathbf{p}, \mathbf{r} \rangle} \right) &\leqslant \langle \mathbf{q}, \mathbf{r} \rangle \ln \left(\frac{\langle \mathbf{q}, \mathbf{r} \rangle}{\langle \mathbf{p}, \mathbf{r} \rangle} \right) + \min_{1 \leqslant i \leqslant n} \{p_i r_i\} \bar{S}(\mathbf{q}, \mathbf{p}) \\ &\leqslant D(\mathbf{q}, \mathbf{p}; \mathbf{r}) \\ &\leqslant \langle \mathbf{q}, \mathbf{r} \rangle \ln \left(\frac{\langle \mathbf{q}, \mathbf{r} \rangle}{\langle \mathbf{p}, \mathbf{r} \rangle} \right) + \max_{1 \leqslant i \leqslant n} \{p_i r_i\} \bar{S}(\mathbf{q}, \mathbf{p}),\end{aligned} \tag{9.76}$$

where

$$\bar{S}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{n} \frac{q_i}{p_i} \ln \left(\frac{q_i}{p_i} \right) - \sum_{i=1}^{n} \frac{q_i}{p_i} \ln \left(\frac{1}{n} \sum_{i=1}^{n} \frac{q_i}{p_i} \right) \geq 0.$$

If in addition $\mathbf{r} = \mathbf{e} = (1, \ldots, 1)$, *then*

$$0 \leqslant \min_{1 \leqslant i \leqslant n} \{p_i\} \bar{S}(\mathbf{q}, \mathbf{p}) \leqslant D(\mathbf{q}, \mathbf{p}) \leqslant \max_{1 \leqslant i \leqslant n} \{p_i\} \bar{S}(\mathbf{q}, \mathbf{p}). \tag{9.77}$$

Proof. Applying (9.66) to the convex function $f(t) = t \ln t$, $t > 0$, we get

$$\begin{aligned}0 &\leqslant \min_{1 \leqslant i \leqslant n} \{p_i r_i\} \bar{S}(\mathbf{q}, \mathbf{p}) \\ &\leqslant D(\mathbf{q}, \mathbf{p}; \mathbf{r}) - \langle \mathbf{q}, \mathbf{r} \rangle \ln \left(\frac{\langle \mathbf{q}, \mathbf{r} \rangle}{\langle \mathbf{p}, \mathbf{r} \rangle} \right) \leqslant \max_{1 \leqslant i \leqslant n} \{p_i r_i\} \bar{S}(\mathbf{q}, \mathbf{p}),\end{aligned}$$

where $\bar{S}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{n} \frac{q_i}{p_i} \ln \left(\frac{q_i}{p_i} \right) - \sum_{i=1}^{n} \frac{q_i}{p_i} \ln \left(\frac{1}{n} \sum_{i=1}^{n} \frac{q_i}{p_i} \right) \geq 0$, which is equivalent to (9.76). Further, choosing $\mathbf{r} = \mathbf{e} = (1, \ldots, 1)$ we have

$$\langle \mathbf{p}, \mathbf{r} \rangle = \langle \mathbf{p}, \mathbf{e} \rangle = \sum_{i=1}^{n} p_i = 1, \quad \langle \mathbf{q}, \mathbf{r} \rangle = \langle \mathbf{q}, \mathbf{e} \rangle = \sum_{i=1}^{n} q_i = 1,$$

i.e. (9.76) reduces to (9.77). □

9.10 Bounds for the Zipf-Mandelbrot entropy

Using the Zipf-Mandelbrot law we derive some results that include the Zipf-Mandelbrot entropies.

Theorem 9.9 *Let $n \in \mathbb{N}$, $q \geq 0$, $s > 0$, $\mathbf{r} = (r_1, \ldots, r_n) \in [0,\infty)^n$ and $\mathbf{q} = (q_1, \ldots, q_n) \in (0,\infty)^n$. Let $H_{n,q,s}$, $Z(H,q,s)$, $H_{n,q,s,\mathbf{r}}$ and $Z(H,q,s,\mathbf{r})$ be defined by (9.17)-(9.19), respectively. Then*

$$
\begin{aligned}
&\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \frac{1}{q_i} + \sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \left(\frac{\sum_{i=1}^{n} r_i q_i}{\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}}} \right) \\
&- \max_{1 \leqslant i \leqslant n} \left\{ \frac{r_i}{(i+q)^s H_{n,p,s,\mathbf{r}}} \right\} \bar{S}(n,q,s,\mathbf{q},\mathbf{r}) \\
&\leqslant Z(H,q,s,\mathbf{r}) \\
&\leqslant \sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \frac{1}{q_i} + \sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \left(\frac{\sum_{i=1}^{n} r_i q_i}{\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}}} \right) \\
&- \min_{1 \leqslant i \leqslant n} \left\{ \frac{r_i}{(i+q)^s H_{n,p,s,\mathbf{r}}} \right\} \bar{S}(n,q,s,\mathbf{q},\mathbf{r}) \\
&\leqslant \sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \frac{1}{q_i} + \sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \left(\frac{\sum_{i=1}^{n} r_i q_i}{\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}}} \right),
\end{aligned}
\tag{9.78}
$$

where

$$
\bar{S}(n,q,s,\mathbf{q},\mathbf{r}) = n \ln \left(\frac{1}{n} \sum_{i=1}^{n} \frac{q_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \right) - \sum_{i=1}^{n} \ln \left(\frac{q_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \right) \geq 0.
$$

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

$$
\begin{aligned}
&\sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,p,s}} \ln \frac{1}{q_i} + \ln \lambda - \max_{1 \leqslant i \leqslant n} \left\{ \frac{1}{(i+q)^s H_{n,p,s}} \right\} \bar{S}(n,q,s,\mathbf{q}) \\
&\leqslant Z(H,q,s) \\
&\leqslant \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,p,s}} \ln \frac{1}{q_i} + \ln \lambda - \min_{1 \leqslant i \leqslant n} \left\{ \frac{1}{(i+q)^s H_{n,p,s}} \right\} \bar{S}(n,q,s,\mathbf{q})
\end{aligned}
\tag{9.79}
$$

$$\leqslant \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,p,s}} \ln \frac{1}{q_i} + \ln \lambda,$$

where

$$\bar{S}(n,q,s,\mathbf{q}) = n \ln\left(\frac{1}{n}\sum_{i=1}^{n} \frac{q_i}{(i+q)^s H_{n,q,s}}\right) - \sum_{i=1}^{n} \ln\left(\frac{q_i}{(i+q)^s H_{n,q,s}}\right) \geq 0.$$

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

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

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

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

becomes

$$\begin{aligned}
&-\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \frac{1}{(i+q)^s H_{n,q,s,\mathbf{r}}} \\
&= \sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln (i+q)^s H_{n,q,s,\mathbf{r}} \\
&= \sum_{i=1}^{n} \frac{r_i \ln(i+q)^s + r_i \ln H_{n,q,s,\mathbf{r}}}{(i+q)^s H_{n,q,s,\mathbf{r}}} = \sum_{i=1}^{n} \frac{r_i \ln(i+q)^s}{(i+q)^s H_{n,q,s,\mathbf{r}}} + \sum_{i=1}^{n} \frac{r_i \ln H_{n,q,s,\mathbf{r}}}{(i+q)^s H_{n,q,s,\mathbf{r}}} \\
&= \frac{s}{H_{n,q,s,\mathbf{r}}} \sum_{i=1}^{n} \frac{r_i \ln(i+q)}{(i+q)^s} + \frac{\ln H_{n,q,s,\mathbf{r}}}{H_{n,q,s,\mathbf{r}}} \sum_{i=1}^{n} \frac{r_i}{(i+q)^s} = \frac{s}{H_{n,q,s,\mathbf{r}}} \sum_{i=1}^{n} \frac{r_i \ln(i+q)}{(i+q)^s} + \ln H_{n,q,s,\mathbf{r}} \\
&= Z(H,q,s,\mathbf{r}).
\end{aligned}$$

Now applying (9.71) we get required result.
Specially, if we choose $\mathbf{q} = \mathbf{e} = (1,\ldots,1)$, then (9.78) reduces to (9.79). □

Corollary 9.20 *Let $n \in \mathbb{N}$, $q \geq 0$, $s > 0$ and $\mathbf{r} = (r_1,\ldots,r_n) \in [0,\infty)^n$ with $R_n = \sum_{i=1}^{n} r_i$. Let $H_{n,q,s}$, $Z(H,q,s)$, $H_{n,q,s,\mathbf{r}}$ and $Z(H,q,s,\mathbf{r})$ be defined by (9.17)-(9.19), respectively. Then*

$$\begin{aligned}
&\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln\left(\frac{R_n}{\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}}}\right) - \max_{1 \leqslant i \leqslant n}\left\{\frac{r_i}{(i+q)^s H_{n,p,s,\mathbf{r}}}\right\} \bar{S}(n,q,s,\mathbf{r}) \qquad (9.80) \\
&\leqslant Z(H,q,s,\mathbf{r})
\end{aligned}$$

$$\leqslant \sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \left(\frac{R_n}{\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}}} \right) - \min_{1 \leqslant i \leqslant n} \left\{ \frac{r_i}{(i+q)^s H_{n,p,s,\mathbf{r}}} \right\} \bar{S}(n,q,s,\mathbf{r})$$

$$\leqslant \sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}} \ln \left(\frac{R_n}{\sum_{i=1}^{n} \frac{r_i}{(i+q)^s H_{n,q,s,\mathbf{r}}}} \right),$$

where

$$\bar{S}(n,q,s,\mathbf{r}) = n \ln \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s,\mathbf{r}}} \right) - \sum_{i=1}^{n} \ln \left(\frac{1}{(i+q)^s H_{n,q,s,\mathbf{r}}} \right) \geq 0.$$

If in addition $\mathbf{r} = \mathbf{e} = (1,\ldots,1)$, *then*

$$\ln n - \max_{1 \leqslant i \leqslant n} \left\{ \frac{1}{(i+q)^s H_{n,p,s}} \right\} \bar{S}(n,q,s) \leqslant Z(H,q,s) \tag{9.81}$$

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

where

$$\bar{S}(n,q,s) = n \ln \left(\frac{1}{n} \sum_{i=1}^{n} \frac{1}{(i+q)^s H_{n,q,s}} \right) - \sum_{i=1}^{n} \ln \left(\frac{1}{(i+q)^s H_{n,q,s}} \right) \geq 0.$$

Proof. Taking $\mathbf{q} = \mathbf{e} = (1,\ldots,1)$ in (9.78) we get (9.80). Specially, if we choose $\mathbf{r} = \mathbf{e} = (1,\ldots,1)$, then (9.80) reduces to (9.81).ds. □

Bibliography

[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] M. Belis, S. Guia**c**su, *A quantitative-qualitative measure of information in cybernetic systems*, IEEE Trans. Information Theory 14 (1968), 593-594.

[4] 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.

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

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorem for Discrete Memoryless Systems*, Academic Press, New York, 1981

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

[8] L. Egghe and R. Rousseau, *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science*, Elsevier Science Publishers, New York, 1990.

[9] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities*, Cambridge University Press, 2nd ed., Cambridge 1952.

[10] 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

[11] A. Barbir, S. Ivelić Bradanović, Đ. Pečarić, J. Pečarić, *Converse to the Sherman Inequality with applications,* submitted

[12] S. Ivelić Bradanović, N. Latif, J. Pečarić, *On an upper bound for Sherman's inequality.* J. Inequal. Appl. **2016** (2016)

[13] S. Ivelić Bradanović, N. Latif, Đ. Pečarić, J. Pečarić, *Sherman's and related inequalities with applications in information theory, J. Inequal. Appl.* **2018** (2018)

[14] S. Ivelić Bradanović, Đ. Pečarić, J. Pečarić, *New lower and upper bounds for f-divergences with applications in Information Theory,* submitted

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

[16] S. Ivelić Bradanović, J. Pečarić, *Generalizations of Sherman's inequality,* Per. Math. Hung. **74** (2) 2017

[17] J.H. Justice, Maximum Entropy and Bayssian Methods in Applied Statistics, Cambridge University Press, Cambridge, 1986.

[18] 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.

[19] S. Kullback, Information Theory and Statistics, J. Wiley, New York, 1959

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

[21] 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

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

[23] A. W. Marshall, I. Olkin and Barry C. Arnold, *Inequalities: Theory of Majorization and Its Applications (Second Edition)*, Springer Series in Statistics, New York 2011.

[24] M. Mohan, J. Mitter, *On bounds of 'useful' information measure*, Indian J. Pure Appl. Math. **9** (9), 960-964.

[25] M. A. Montemurro, *Beyond the Zipf-Mandelbrot law in quantitative linguistics*. URL: arXiv:cond-mat/0104066v2, (2001).

[26] D. Mouillot, A. Lepretre, *Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity*, Environmental Monitoring and Assessment, Springer, **63** (2), 2000, p. 279-295.

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

[28] J. E. Pečarić, F. Proshan and Y. L. Tong, *Convex Functions, Partial Orderings and Statistical Applications*, Academic Press, Inc. (1992)

[29] 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.

[30] I. Schur, *Uber eine Klasse von Mittelbildungen mit Anwendungen auf die Determinanten- theorie*, Sitzungsber. Berl. Math. Ges. 22 (1923) 9–20

[31] I. Schur, *Collected Works*, Vol. II, A. Brauer, A. Rohrbach (eds.), Springer, Berlin (1973) 418–427

[32] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal 27 (1948) 379 - 423.

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

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

[35] Z. K. Silagadze, *Citations and the Zipf-Mandelbrot Law*, Complex Systems, (1997) **11**, p. 487-499.

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

Chapter 10

Jensen-type inequalities for generalized f-divergences and Zipf-Mandelbrot law

Đilda Pečarić, Josip Pečarić, Dora Pokaz and Mirna Rodić

Abstract. Here we present different inequalities of Jensen-type for generalized f-divergences such as generalized Csiszár f-divergence, generalized Kullbach-Leibler, Hellinger, Rényi and χ^2 divergence, as well as for the generalized Shannon entropy. The applications on the Zipf-Mandelbrot law, which is one specific kind of probability distributions, are also presented.

10.1 Introduction

For a function $f\colon \mathbb{R}_+ \to \mathbb{R}$ and two positive probability distributions $\mathbf{p} = (p_1,\ldots,p_n) \in \mathbb{R}^n_+$, $\mathbf{q} = (q_1,\ldots,q_n) \in \mathbb{R}^n_+$, I. Csiszár in [6] introduced the f-divergence functional by

$$C_f(\mathbf{q},\mathbf{p}) = \sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right). \tag{10.1}$$

I. Csiszár studied (10.1) under assumption that function f is convex. Independently, Morimoto [18] and Ali and Silvey [1] also introduced and studied these divergences. Still, (10.1) is widely known as Csiszár f-divergence. These divergences are well known in probability theory, in information theory, in statistical physics, economics, biology, etc. In probability theory, an f-divergence is a function $Df(P \parallel Q)$ that measures the difference between two probability distributions P and Q. Intuitively, the divergence is an average, weighted by the function f, of the odds ratio given by P and Q. There are lots of articles on that subject, both recent and older such as [4], [5], [10], [11], [12], [15] and [16].

In one part of our work, we are following the idea of Y. J. Cho, M. Matić, and J. Pečarić [3], but in discrete case and additionally generalized. In that way we get Jensen's type inequalities for Lipschitzian functions in terms of generalized Csiszár's functional (see [19]). We recall that a real-valued function $f\colon \mathbb{R} \to \mathbb{R}$ is called Lipschitz continuous if there exists a positive real constant L such that, for all $x_1, x_2 \in \mathbb{R}$

$$|f(x_1) - f(x_2)| \le L|x_1 - x_2|$$

holds. Shortly, we call those functions L-Lipschitzian or just Lipschitzian. We also go through some of the most frequent types of f-divergences. Namely, we state Jensen's type inequality for Lipschitzian functions involving the Kullbach-Leibler divergence, the Hellinger divergence, the Rényi divergence and χ^2-divergence, all generalized.

As the Jensen inequality is important in obtaining inequalities for divergences between probability distributions, there are many papers dealing with inequalities for divergences and entropies (see for example [11], [17] or [20]). By means of one Jensen-type inequality which is characterized via several different Green functions, we will here also derive some new inequalities for divergences (see [21]).

At the end of our results, we will also give the applications on the Zipf-Mandelbrot law, as one specific kind of probability distributions. The results given here are presented in [19] and [21].

10.2 Preliminary results

The discrete Jensen inequality states that

$$\varphi\left(\frac{1}{U_n}\sum_{i=1}^{n} u_i x_i\right) \le \frac{1}{U_n}\sum_{i=1}^{n} u_i \varphi(x_i) \tag{10.2}$$

holds for a convex function $\varphi : I \to \mathbf{R}$, $I \subseteq \mathbf{R}$, an n-tuple $\mathbf{x} = (x_1, \dots, x_n)$ $(n \ge 2)$ and nonnegative n-tuple $\mathbf{u} = (u_1, \dots, u_n)$, such that $U_n = \sum_{i=1}^n u_i > 0$. In order to simplify the notation here we shall use the common notation: $U_n = \sum_{i=1}^n u_i$ and $\overline{x} = \frac{1}{U_n}\sum_{i=1}^n u_i x_i$.

In [22] and [23] we have the generalization of that result. Namely, there is also allowed that u_i are negative with their sum different from 0, but we have a supplementary demand on u_i, x_i using the Green functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$, $(k = 0, 1, 2, 3, 4)$ defined by

$$G_0(t,s) = \begin{cases} \frac{(t-\beta)(s-\alpha)}{\beta-\alpha} & \text{for } \alpha \le s \le t, \\ \frac{(s-\beta)(t-\alpha)}{\beta-\alpha} & \text{for } t \le s \le \beta. \end{cases} \tag{10.3}$$

$$G_1(t,s) = \begin{cases} \alpha - s, & \text{for } \alpha \le s \le t, \\ \alpha - t, & \text{for } t \le s \le \beta. \end{cases} \tag{10.4}$$

$$G_2(t,s) = \begin{cases} t - \beta, & \text{for } \alpha \le s \le t, \\ s - \beta, & \text{for } t \le s \le \beta. \end{cases} \tag{10.5}$$

$$G_3(t,s) = \begin{cases} t - \alpha, & \text{for } \alpha \le s \le t, \\ s - \alpha, & \text{for } t \le s \le \beta. \end{cases} \tag{10.6}$$

$$G_4(t,s) = \begin{cases} \beta - s, & \text{for } \alpha \le s \le t, \\ \beta - t, & \text{for } t \le s \le \beta. \end{cases} \tag{10.7}$$

The following result holds true:

Theorem 10.1 *Let $x_i \in [a,b] \subseteq [\alpha,\beta]$, $u_i \in \mathbb{R}$ $(i = 1,\dots,n)$, be such that $U_n \neq 0$ and $\overline{x} \in [\alpha,\beta]$, and let $\varphi : [\alpha,\beta] \to \mathbb{R}$, $\varphi \in C^2([\alpha,\beta])$. Let the functions $G_k : [\alpha,\beta]\times[\alpha,\beta] \to \mathbb{R}$ $(k = 0,1,2,3,4)$ be as defined in* (10.3)-(10.7). *Furthermore, let $p,q \in \mathbb{R}$, $1 \le p,q \le \infty$, be such that $\frac{1}{p}+\frac{1}{q} = 1$. Then*

$$\left| \frac{1}{U_n}\sum_{i=1}^{n} u_i\varphi(x_i) - \varphi(\overline{x}) \right| \le Q \cdot \left\|\varphi''\right\|_p \tag{10.8}$$

holds, where

$$Q = \begin{cases} \left[\int_\alpha^\beta \left| \frac{1}{U_n}\sum_{i=1}^n u_i G_k(x_i,s) - G_k(\overline{x},s) \right|^q ds \right]^{\frac{1}{q}} & \textit{for } q \neq \infty; \\ \sup_{s\in[\alpha,\beta]} \left\{ \left| \frac{1}{U_n}\sum_{i=1}^n u_i G_k(x_i,s) - G_k(\overline{x},s) \right| \right\} & \textit{for } q = \infty. \end{cases} \tag{10.9}$$

Proof. Using the functions G_k $(k = 0,1,2,3,4)$, every function $\varphi : [\alpha,\beta] \to \mathbb{R}$, $\varphi \in C^2([\alpha,\beta])$, can be represented as

$$\varphi(x) = \frac{\beta - x}{\beta - \alpha}\varphi(\alpha) + \frac{x-\alpha}{\beta-\alpha}\varphi(\beta) + \int_\alpha^\beta G_0(x,s)\varphi''(s)ds \tag{10.10}$$

$$\varphi(x) = \varphi(\alpha) + (x-\alpha)\varphi'(\beta) + \int_\alpha^\beta G_1(x,s)\varphi''(s)ds, \tag{10.11}$$

$$\varphi(x) = \varphi(\beta) + (x-\beta)\varphi'(\alpha) + \int_\alpha^\beta G_2(x,s)\varphi''(s)ds, \tag{10.12}$$

$$\varphi(x) = \varphi(\beta) - (\beta-\alpha)\varphi'(\beta) + (x-\alpha)\varphi'(\alpha) + \int_\alpha^\beta G_3(x,s)\varphi''(s)ds, \tag{10.13}$$

$$\varphi(x) = \varphi(\alpha) + (\beta-\alpha)\varphi'(\alpha) - (\beta-x)\varphi'(\beta) + \int_\alpha^\beta G_4(x,s)\varphi''(s)ds, \tag{10.14}$$

which can be easily shown by integrating by parts. It is also easy to show by some calculation that for every such function φ and for any $k \in \{0,1,2,3,4\}$ it holds:

$$\frac{1}{U_n}\sum_{i=1}^{n} u_i\varphi(x_i) - \varphi(\overline{x}) = \int_{\alpha}^{\beta}\left(\frac{1}{U_n}\sum_{i=1}^{n} u_i G_k(x_i,s) - G_k(\overline{x},s)\right)\varphi''(s)ds. \tag{10.15}$$

Using the triangle inequality for integrals and then applying the Hölder inequality, we get the following:

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

and we get the result given in our theorem. □

10.3 Jensen's type inequalities for the generalized Csiszár f-divergence

The definition of the Csiszár f-divergence functional given in (10.1) can be further generalized using weights. For a function $f\colon \mathbb{R}_+ \to \mathbb{R}$ and $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathbb{R}_+^n$, we define the generalized Csiszár f-divergence by

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right). \tag{10.16}$$

In order to simplify our results, we introduce the following notations

$$P_r = \sum_{i=1}^{n} r_i p_i, \tag{10.17}$$

$$\overline{Q}_r = \frac{1}{P_r}\sum_{i=1}^{n} r_i q_i. \tag{10.18}$$

Our first result of this section is the following Jensen's type inequality for Lipschitzian function based on the idea of Y. J. Cho et al. [3].

Theorem 10.2 *Suppose $p_i, q_i, r_i\,(i \in \mathbb{N})$ are positive real numbers. If $f: \mathbb{R}_+ \to \mathbb{R}$ is an L-Lipschitzian function, then*

$$\left| \frac{1}{P_r} C_f(\boldsymbol{q}, \mathbf{p}; \boldsymbol{r}) - f\left(\overline{Q}_r\right) \right| \leq \frac{L}{P_r} \sum_{i=1}^{n} r_i p_i \left| \frac{q_i}{p_i} - \overline{Q}_r \right| \tag{10.19}$$

holds, where $C_f(\boldsymbol{q}, \mathbf{p}; \boldsymbol{r})$, P_r and $\overline{Q}_r$ are defined by (10.16), (10.17) *and* (10.18) *respectively.*

Proof. The inequality (10.19) follows by elementary techniques

$$\begin{aligned} &\left| \frac{1}{P_r} \sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right) - f\left(\frac{1}{P_r} \sum_{i=1}^{n} r_i q_i\right) \right| = \frac{1}{P_r} \left| \sum_{i=1}^{n} r_i p_i \left[f\left(\frac{q_i}{p_i}\right) - f(\overline{Q}_r) \right] \right| \\ &\leq \frac{1}{P_r} \sum_{i=1}^{n} r_i p_i \left| f\left(\frac{q_i}{p_i}\right) - f(\overline{Q}_r) \right| \leq \frac{L}{P_r} \sum_{i=1}^{n} r_i p_i \left| \frac{q_i}{p_i} - \overline{Q}_r \right| \end{aligned}$$

□

The following inequality for bounded sequence $(q_1, \ldots, q_n)$ is also based on [3] .

Theorem 10.3 *Let $f: \mathbb{R}_+ \to \mathbb{R}$ be an L-Lipschitzian function, let $p_i, r_i\,(i \in \mathbb{N})$ be positive real numbers such that C_f, P_r and $\overline{Q}_r$ are defined by* (10.16), (10.17) *and* (10.18). *If there exist $m, M \in \mathbb{R}$ such that $mp_i \leq q_i \leq Mp_i \;\; (i \in \mathbb{N})$, then*

$$\begin{aligned} &\left| \frac{M - \overline{Q}_r}{M - m} f(m) + \frac{\overline{Q}_r - m}{M - m} f(M) - \frac{1}{P_r} C_f(\boldsymbol{q}, \mathbf{p}; \boldsymbol{r}) \right| \\ &\leq \frac{2L}{P_r(M - m)} \sum_{i=1}^{n} r_i p_i \left(M - \frac{q_i}{p_i} \right) \left(\frac{q_i}{p_i} - m \right). \end{aligned} \tag{10.20}$$

holds.

Proof. Starting from the left-hand side of (10.20), we get

$$\begin{aligned} &\left| \frac{M - \overline{Q}_r}{M - m} f(m) + \frac{\overline{Q}_r - m}{M - m} f(M) - \frac{1}{P_r} \sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right) \right| \\ &= \frac{1}{P_r} \left| \sum_{i=1}^{n} r_i p_i \left[\frac{M - \frac{q_i}{p_i}}{M - m} f(m) + \frac{\frac{q_i}{p_i} - m}{M - m} f(M) - f\left(\frac{q_i}{p_i}\right) \right] \right| \\ &\leq \frac{1}{P_r} \sum_{i=1}^{n} r_i p_i \left| \frac{M - \frac{q_i}{p_i}}{M - m} f(m) + \frac{\frac{q_i}{p_i} - m}{M - m} f(M) - f\left(\frac{q_i}{p_i}\right) \right| \\ &= \frac{1}{P_r} \sum_{i=1}^{n} r_i p_i \left| \frac{M - \frac{q_i}{p_i}}{M - m} \left(f(m) - f\left(\frac{q_i}{p_i}\right) \right) + \frac{\frac{q_i}{p_i} - m}{M - m} \left(f(M) - f\left(\frac{q_i}{p_i}\right) \right) \right| \\ &\leq \frac{1}{P_r} \sum_{i=1}^{n} r_i p_i \left[\frac{M - \frac{q_i}{p_i}}{M - m} \left| f(m) - f\left(\frac{q_i}{p_i}\right) \right| + \frac{\frac{q_i}{p_i} - m}{M - m} \left| f(M) - f\left(\frac{q_i}{p_i}\right) \right| \right] \end{aligned}$$

$$\leq \frac{2L}{P_r(M-m)}\sum_{i=1}^{n} r_i p_i \left(M - \frac{q_i}{p_i}\right)\left(\frac{q_i}{p_i} - m\right)$$

using the properties of the absolute value function. □

For the following result, we apply Theorem 10.1 on $C_f(\mathbf{q},\mathbf{p};\mathbf{r})$.

Theorem 10.4 *Let the functions $G_k : [\alpha,\beta]\times[\alpha,\beta] \to \mathbb{R}$ $(k = 0,1,2,3,4)$ be as defined in* (10.3)-(10.7). *Let $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathbb{R}^n_+$ be such that*

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

where $\overline{Q}_r$ is as defined in (10.18). Furthermore, let $p,q \in \mathbb{R}$, $1 \leq p,q \leq \infty$, be such that $\frac{1}{p}+\frac{1}{q} = 1$.

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

$$\left|\frac{1}{P_r}C_f(\mathbf{q},\mathbf{p};\mathbf{r}) - f\left(\overline{Q}_r\right)\right| \leq Q\cdot\left\|f''\right\|_p \tag{10.21}$$

holds, where P_r and $C_f(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (10.17) and (10.16) respectively, and

$$Q = \begin{cases} \left[\int_\alpha^\beta \left|\frac{1}{P_r}\sum_{i=1}^n r_i p_i G_k\left(\frac{q_i}{p_i},s\right) - G_k\left(\overline{Q}_r,s\right)\right|^q ds\right]^{\frac{1}{q}}, & \text{for } q \neq \infty; \\ \sup_{s\in[\alpha,\beta]}\left\{\left|\frac{1}{P_r}\sum_{i=1}^n r_i p_i G_k\left(\frac{q_i}{p_i},s\right) - G_k\left(\overline{Q}_r,s\right)\right|\right\}, & \text{for } q = \infty. \end{cases} \tag{10.22}$$

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

$$\left|\frac{1}{P_r}C_{id\cdot f}(\mathbf{q},\mathbf{p};\mathbf{r}) - \overline{Q}_r\cdot f\left(\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i}\right)\right| \leq Q\cdot\left\|(id\cdot f)''\right\|_p \tag{10.23}$$

holds, where id is the identity function, $C_{id\cdot f}(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i q_i f\left(\frac{q_i}{p_i}\right)$ and Q is as defined in (10.22).

Proof.

(a) The result follows directly from Theorem 10.1 by substitution $\varphi := f$,

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

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

□

10.4 Inequalities for different types of generalized $f-$divergences

Now, we will consider some of the most important examples of f-divergences.

The Kullback-Leibler divergence (see [13], [14]) for $\mathbf{p},\mathbf{q} \in \mathbb{R}_+^n$ is defined by

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

It is easy to see that the Kullback-Leibler divergence is in fact the Csiszár f-divergence, where $f(t) = t\log t, t > 0$. We can generalize this f-divergence, and we define the generalized Kullback-Leibler divergence by

$$KL(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i q_i \log\frac{q_i}{p_i}, \tag{10.24}$$

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

Proposition 10.1 *Let $\mathbf{p},\boldsymbol{q},\boldsymbol{r} \in \mathbb{R}_+^n$, P_r, $\overline{Q}_r$ and $KL(\boldsymbol{q},\mathbf{p};\boldsymbol{r})$ are defined by* (10.17), (10.18) *and* (10.24). *If there exist $m,M \in \mathbb{R}_+$ such that $mp_i \le q_i \le Mp_i$, $i \in \mathbb{N}$, then inequalities*

$$\begin{aligned}&\left|KL(\mathbf{q},\mathbf{p};\mathbf{r}) - \sum_{i=n}^{n} r_i q_i \log\frac{\sum_{i=n}^{n} r_i q_i}{P_r}\right| \\ &\le \max\{|\log m + 1|, |\log M + 1|\} \sum_{i=n}^{n} r_i p_i \left|\frac{q_i}{p_i} - \overline{Q_r}\right|\end{aligned} \tag{10.25}$$

and

$$\begin{aligned}&\left|\frac{M-\overline{Q}_r}{M-m} m\log m + \frac{\overline{Q}_r - m}{M-m} M\log M - \frac{1}{P_r} KL(\boldsymbol{q},\mathbf{p};\boldsymbol{r})\right| \\ &\le \max\{|\log m + 1|, |\log M + 1|\} \frac{2}{P_r(M-m)} \sum_{i=1}^{n} r_i p_i \left(M - \frac{q_i}{p_i}\right)\left(\frac{q_i}{p_i} - m\right)\end{aligned} \tag{10.26}$$

hold.

Proof. The inequalities (10.25) and (10.26) are derived from (10.19) and (10.20) for $f(t) = t\log t, t > 0$. In this case $C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i p_i \frac{q_i}{p_i} \log\frac{q_i}{p_i} = KL(\mathbf{q},\mathbf{p};\mathbf{r})$ and $L = \sup_{t\in[m,M]} |\log t + 1|$, since $f'(t) = \log t + 1$ is bounded on $[m,M]$. □

Proposition 10.2 *Let the functions $G_k : [\alpha,\beta] \times [\alpha,\beta] \to \mathbb{R}$ $(k = 0,1,2,3,4)$ be as defined in* (10.3)-(10.7). *Let $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathbb{R}_+^n$ be such that*

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

where $\overline{Q}_r$ is as defined in (10.18). Furthermore, let $p,q \in \mathbb{R}$, $1 \le p,q \le \infty$, be such that $\frac{1}{p}+\frac{1}{q}=1$.

Then

$$\left|\frac{1}{P_r}KL(\mathbf{q},\mathbf{p};\mathbf{r}) - \overline{Q}_r \cdot \log\left(\frac{\sum_{i=1}^n q_i}{\sum_{i=1}^n p_i}\right)\right| \le Q \cdot \left\|(id\cdot \log)''\right\|_p \tag{10.27}$$

holds, where P_r is as defined in (10.17), id is the identity function and Q is as defined in (10.22).

Proof. The result follows from Theorem 10.4 (b) by substitution $f := \log$ (i.e. from Theorem 10.4 (a) by substitution $f(t) := t\log t$, $t > 0$). □

For $\mathbf{p},\mathbf{q} \in \mathbb{R}_+^n$, the Hellinger divergence (as given in [9])

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

is the Csiszár f-divergence for $f(t) = (1-\sqrt{t})^2$, $t > 0$. As before, we also generalize this divergence for $\mathbf{r} \in \mathbb{R}_+^n$ with

$$He(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^n r_i(\sqrt{q_i} - \sqrt{p_i})^2. \tag{10.28}$$

We have the following estimations.

Proposition 10.3 *Let $\mathbf{p},\boldsymbol{q},\boldsymbol{r} \in \mathbb{R}_+^n$, $P_r, \overline{Q}_r$ and $He(\boldsymbol{q},\mathbf{p};\boldsymbol{r})$ be defined by* (10.17), (10.18) *and* (10.28). *If there exist $m,M \in \mathbb{R}_+$ such that $mp_i \le q_i \le Mp_i$, $i \in \mathbb{N}$, then inequalities*

$$\left|He(\mathbf{q},\mathbf{p};\mathbf{r}) - \left(\sqrt{P_r} - \sqrt{\sum_{i=n}^n r_i q_i}\right)^2\right|$$
$$\le \max\left\{\frac{|m-\sqrt{m}|}{m}, \frac{|M-\sqrt{M}|}{M}\right\} \sum_{i=n}^n r_i p_i \left|\frac{q_i}{p_i} - \overline{Q}_r\right| \tag{10.29}$$

and

$$\left|\frac{M-\overline{Q}_r}{M-m}(1-\sqrt{m})^2 + \frac{\overline{Q}_r - m}{M-m}(1-\sqrt{M})^2 - \frac{1}{P_r}He(\boldsymbol{q},\mathbf{p};\boldsymbol{r})\right| \tag{10.30}$$
$$\le \frac{2}{P_r(M-m)} \max\left\{\frac{|m-\sqrt{m}|}{m}, \frac{|M-\sqrt{M}|}{M}\right\} \sum_{i=1}^n r_i p_i \left(M - \frac{q_i}{p_i}\right)\left(\frac{q_i}{p_i} - m\right)$$

hold.

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

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

and

$$L = \sup_{t\in[m,M]} \left|1-\frac{1}{\sqrt{t}}\right| = \max\left\{\left|1-\frac{1}{\sqrt{m}}\right|, \left|1-\frac{1}{\sqrt{M}}\right|\right\}$$
$$= \max\left\{\frac{|m-\sqrt{m}|}{m}, \frac{|M-\sqrt{M}|}{M}\right\}.$$

So, inequalities (10.29) and (10.30) follow from (10.19) and (10.20). □

Proposition 10.4 *Let the functions* $G_k : [\alpha,\beta]\times[\alpha,\beta]\to\mathbb{R}$ $(k=0,1,2,3,4)$ *be as defined in* (10.3)-(10.7). *Let* $\mathbf{p},\mathbf{q},\mathbf{r}\in\mathbb{R}^n_+$ *be such that*

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

where $\overline{Q}_r$ *is as defined in (10.18). Furthermore, let* $p,q\in\mathbb{R}$, $1\le p,q\le\infty$, *be such that* $\frac{1}{p}+\frac{1}{q}=1$.

Then

$$\left|\frac{1}{P_r}He(\mathbf{q},\mathbf{p};\mathbf{r}) - \left(1-\sqrt{\overline{Q}_r}\right)^2\right| \le Q\cdot\left\|f''\right\|_p \tag{10.31}$$

holds, where P_r *is as defined in (10.17),* $f(t)=(1-\sqrt{t})^2$, $t>0$, *and* Q *is as defined in (10.22).*

Proof. The result follows from Theorem 10.4 (a) by substitution $f(t)=(1-\sqrt{t})^2$, $t>0$. □

The γ-order entropy known as the Rényi divergence ([24]) is defined by

$$Re_\gamma(\mathbf{q},\mathbf{p}) = \sum_{i=1}^n p_i^{1-\gamma}q_i^\gamma,\ \text{ for } \mathbf{p},\mathbf{q}\in\mathbb{R}^n_+,\ \gamma\in\langle 1,+\infty\rangle. \tag{10.32}$$

We generalize (10.32) by

$$Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^n r_i p_i^{1-\gamma}q_i^\gamma,\ r\in\mathbb{R}^n_+. \tag{10.33}$$

For this generalized entropy we have the following results.

Proposition 10.5 *Let* $\mathbf{p},\mathbf{q},\mathbf{r}\in\mathbb{R}^n_+$, P_r, $\overline{Q}_r$ *and* $Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r})$ *be defined by* (10.17), (10.18) *and* (10.33) . *If there exist* $m,M\in\mathbb{R}_+$ *such that* $mp_i\le q_i\le Mp_i$, $i\in\mathbb{N}$, *then inequalities*

$$\left|Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r}) - P_r^{1-\gamma}\left(\sum_{i=1}^n r_iq_i\right)^\gamma\right| \le \gamma M^{\gamma-1}\sum_{i=1}^n r_ip_i\left|\frac{q_i}{p_i}-\overline{Q}_r\right| \tag{10.34}$$

and

$$\left|\frac{M-\overline{Q}_r}{M-m}m^\gamma + \frac{\overline{Q}_r-m}{M-m}M^\gamma - \frac{1}{P_r}Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r})\right| \tag{10.35}$$

$$\leq \frac{2\gamma M^{\gamma-1}}{P_r(M-m)} \sum_{i=1}^{n} r_i p_i \left(M - \frac{q_i}{p_i}\right)\left(\frac{q_i}{p_i} - m\right)$$

hold.

Proof. For $f(t) = t^\gamma, t > 0, \gamma > 1$, we have

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

and

$$L = \sup_{t\in[m,M]} \left|\gamma t^{\gamma-1}\right| = \gamma \sup_{t\in[m,M]} \left|t^{\gamma-1}\right| = \gamma M^{\gamma-1},$$

so we obtain (10.34) and (10.35) from (10.19) and (10.20). □

Proposition 10.6 *Let the functions* $G_k : [\alpha,\beta]\times[\alpha,\beta] \to \mathbb{R}$ $(k = 0,1,2,3,4)$ *be as defined in* (10.3)-(10.7). *Let* $\mathbf{p},\mathbf{q},\mathbf{r} \in \mathbb{R}_+^n$ *be such that*

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

where $\overline{Q}_r$ *is as defined in (10.18). Furthermore, let* $p,q \in \mathbb{R}$, $1 \leq p,q \leq \infty$, *be such that* $\frac{1}{p}+\frac{1}{q} = 1$.

Then

$$\left|\frac{1}{P_r} Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r}) - \overline{Q}_r^\gamma\right| \leq Q \cdot \left\|f''\right\|_p \tag{10.36}$$

holds, where P_r *is as defined in (10.17),* $f(t) = t^\gamma$ $(t > 0,\ \gamma > 1)$, *and* Q *is as defined in (10.22).*

Proof. The result follows from Theorem 10.4 (a) for $f(t) = t^\gamma$ $(t > 0,\ \gamma > 1)$. □

Our next interesting result concerns with the χ^2-divergence defined by

$$D_{\chi^2}(\mathbf{q},\mathbf{p}) = \sum_{i=1}^{n} \frac{(q_i - p_i)^2}{p_i}, \text{ for } \mathbf{p},\mathbf{q} \in \mathbb{R}_+^n.$$

For the generalized χ^2-divergence

$$D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}) = \sum_{i=1}^{n} r_i \frac{(q_i - p_i)^2}{p_i}, \ \mathbf{r} \in \mathbb{R}_+^n \tag{10.37}$$

we have the following results.

Proposition 10.7 *Let* $\mathbf{p},\boldsymbol{q},\boldsymbol{r} \in \mathbb{R}_+^n$, P_r, $\overline{Q}_r$ *and* $D_{\chi^2}(\boldsymbol{q},\mathbf{p};\boldsymbol{r})$ *be defined by* (10.17), (10.18) *and* (10.37). *If there exist* $m,M \in \mathbb{R}$ *such that* $mp_i \leq q_i \leq Mp_i$, $i \in \mathbb{N}$, *then inequalities*

$$\left|D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}) - \frac{1}{P_r}\left(\sum_{i=1}^{n} r_i q_i - P_r\right)^2\right| \leq 2\max\{|m-1|,|M-1|\} \sum_{i=1}^{n} r_i p_i \left|\frac{q_i}{p_i} - \overline{Q}_r\right| \tag{10.38}$$

and

$$\left|\frac{M-\overline{Q}_r}{M-m}(m-1)^2+\frac{\overline{Q}_r-m}{M-m}(M-1)^2-\frac{1}{P_r}D_{\chi^2}(\boldsymbol{q},\mathbf{p};\boldsymbol{r})\right| \tag{10.39}$$
$$\leq \frac{4}{P_r(M-m)}\max\{|m-1|,|M-1|\}\sum_{i=1}^{n} r_i p_i\left(M-\frac{q_i}{p_i}\right)\left(\frac{q_i}{p_i}-m\right)$$

hold.

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

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

Since $f^{'}(t)=2(t-1)$, we have

$$L=2\sup_{t\in[m,M]}|t-1|=2\max\{|m-1|,|M-1|\}.$$

Inequalities (10.38) and (10.39) follow from (10.19) and (10.20). □

Proposition 10.8 *Let the functions* $G_k:[\alpha,\beta]\times[\alpha,\beta]\to\mathbb{R}$ $(k=0,1,2,3,4)$ *be as defined in* (10.3)-(10.7). *Let* $\mathbf{p},\mathbf{q},\mathbf{r}\in\mathbb{R}^n_+$ *be such that*

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

where $\overline{Q}_r$ *is as defined in (10.18). Furthermore, let* $p,q\in\mathbb{R}$, $1\leq p,q\leq\infty$, *be such that* $\frac{1}{p}+\frac{1}{q}=1$.

Then

$$\left|\frac{1}{P_r}D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r})-(\overline{Q}_r-1)^2\right|\leq Q\cdot\left\|f''\right\|_p \tag{10.40}$$

holds, where P_r *is as defined in (10.17),* $f(t)=(t-1)^2$, $t>0$, *and* Q *is as defined in (10.22).*

Proof. The result follows from Theorem 10.4 (a) by substitution $f(t)=(t-1)^2$, $t>0$. □

The Shannon entropy ([11]) of a positive probability distribution $\mathbf{p}=(p_1,\ldots,p_n)$ is defined by

$$H(\mathbf{p})=-\sum_{i=1}^{n} p_i\log(p_i). \tag{10.41}$$

It is easy to see that (10.41) is a special case of (10.1) for $\mathbf{q}=(1,\ldots,1)\in\mathbb{R}^n_+$ and function $f(t)=\log t, t>0$. We can also generalize Shannon entropy with weights $\mathbf{r}\in\mathbb{R}^n_+$ as follows

$$H(\mathbf{p};\mathbf{r})=-\sum_{i=1}^{n} r_i p_i\log(p_i). \tag{10.42}$$

Proposition 10.9 *Let $\mathbf{p}, \boldsymbol{r} \in \mathbb{R}_+^n$ and P_r and $H(\mathbf{p}; \boldsymbol{r})$ be defined by* (10.17) *and* (10.42). *If there exist $m, M \in \mathbb{R}$ such that $m \leq \frac{1}{p_i} \leq M$, $i \in \mathbb{N}$, then inequalities*

$$\left|H(\mathbf{p}; \boldsymbol{r}) - P_r \log(\overline{Q}_r)\right| \leq \frac{1}{m} \sum_{i=1}^{n} r_i p_i \left|\frac{1}{p_i} - \overline{Q}_r\right| \tag{10.43}$$

and

$$\begin{aligned} &\left|\frac{M - \overline{Q}_r}{M - m} f(m) + \frac{\overline{Q}_r - m}{M - m} f(M) - \frac{1}{P_r} H(\mathbf{p}; \boldsymbol{r})\right| \\ &\leq \frac{2}{m(M-m)P_r} \sum_{i=1}^{n} r_i p_i \left(M - \frac{1}{p_i}\right)\left(\frac{1}{p_i} - m\right) \end{aligned} \tag{10.44}$$

hold, where $\overline{Q}_r = \frac{1}{P_r} \sum_{i=1}^{n} r_i$.

Proof. For $f(t) = \log t, t > 0$ and $\mathbf{q} = (1, \ldots, 1)$, we have

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

Since $f'(t) = \frac{1}{t}$, the Lipschitz constant in this case is

$$L = \sup_{t \in [m,M]} \left|\frac{1}{t}\right| = \max\left\{\left|\frac{1}{m}\right|, \left|\frac{1}{M}\right|\right\} = \max\left\{\frac{1}{m}, \frac{1}{M}\right\} = \frac{1}{m}.$$

Inequalities (10.43) and (10.44) are following from (10.19) and (10.20). □

Proposition 10.10 *Let the functions $G_k : [\alpha, \beta] \times [\alpha, \beta] \to \mathbb{R}$ $(k = 0, 1, 2, 3, 4)$ be as defined in* (10.3)-(10.7). *Let $\mathbf{p}, \mathbf{r} \in \mathbb{R}_+^n$ be such that*

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

where P_r is as defined in (10.17). Furthermore, let $p, q \in \mathbb{R}$, $1 \leq p, q \leq \infty$, be such that $\frac{1}{p} + \frac{1}{q} = 1$.
Then

$$\left|\frac{1}{P_r} H(\mathbf{p}; \mathbf{r}) - \log\left(\frac{1}{P_r} \sum_{i=1}^{n} r_i\right)\right| \leq Q \cdot \left\|\log''\right\|_p \tag{10.45}$$

holds, where Q is as defined in (10.22).

Proof. The result follows from Theorem 10.4 (a) by substitution $f := \log$ and $\mathbf{q} = (1, \ldots, 1)$. □

10.5 The mappings of H and F

In this section, we study discrete general case of the mappings called H and F introduced in [7] and [8]. For a given function $f\colon I \subset \mathbb{R} \to \mathbb{R}$ and for $a,b \in I$, $a<b$, S. S. Dragomir considered the following two mappings $H,F\colon [0,1] \to \mathbb{R}$ defined by

$$H(t) = \frac{1}{b-a}\int_a^b f\left(tx+(1-t)\frac{a+b}{2}\right)dx$$

and

$$F(t) = \frac{1}{(b-a)^2}\int_a^b\int_a^b f(tx+(1-t)y)\,dxdy$$

for all $t \in [0,1]$. Under assumption of convexity of f, mappings H and F have been tested on convexity on $[0,1]$, monotonicity and other properties. On this lead, Cho et al. in [3] also studied generalized functions of this type.

In this section, we consider f to be Lipschitzian function and dealing with discrete generalization, so our next results come naturally. We prove some of the properties of the functions F and H, such as Lipschitz property.

Theorem 10.5 *Let $\mathbf{p},\boldsymbol{r} \in \mathbb{R}^n_+$ and $\boldsymbol{q} \in \mathbb{R}^n$ and $f\colon \mathbb{R} \to \mathbb{R}$ be an L-Lipschitzian function. For a mapping $H\colon [0,1] \to \mathbb{R}$ defined by*

$$H(\lambda) = \frac{1}{P_r}\sum_{i=1}^n p_i r_i f\left(\lambda\frac{q_i}{p_i}+(1-\lambda)\overline{Q}_r\right) \tag{10.46}$$

we have the following:

(i) the mapping H is L_1-Lipschitzian on $[0,1]$, where

$$L_1 = \frac{L}{P_r}\sum_{i=1}^n p_i r_i\left(\frac{q_i}{p_i}-\overline{Q}_r\right) \tag{10.47}$$

(ii) the inequalities

$$\left|H(\lambda)-\frac{1}{P_r}\sum_{i=1}^n r_i p_i f\left(\frac{q_i}{p_i}\right)\right| \le (1-\lambda)L_1, \tag{10.48}$$

$$\left|f\left(\frac{1}{P_r}\sum_{i=1}^n r_i q_i\right)-H(\lambda)\right| \le \lambda L_1 \tag{10.49}$$

and

$$\left|H(\lambda)-\frac{\lambda}{P_r}\sum_{i=1}^n r_i p_i f\left(\frac{q_i}{p_i}\right)-(1-\lambda)f(\overline{Q}_r)\right| \le 2\lambda(1-\lambda)L_1 \tag{10.50}$$

hold, for all $\lambda \in [0,1]$.

Proof. For $\lambda_1, \lambda_2 \in [0,1]$, we calculate

$$\begin{aligned}
&|H(\lambda_2) - H(\lambda_1)| \\
&= \frac{1}{P_r}\left|\sum_{i=1}^{n} p_i r_i \left[f\left(\lambda_2 \frac{q_i}{p_i} + (1-\lambda_2)\overline{Q}_r\right) - f\left(\lambda_1 \frac{q_i}{p_i} + (1-\lambda_1)\overline{Q}_r\right)\right]\right| \\
&\leq \frac{1}{P_r}\sum_{i=1}^{n} p_i r_i \left|f\left(\lambda_2 \frac{q_i}{p_i} + (1-\lambda_2)\overline{Q}_r\right) - f\left(\lambda_1 \frac{q_i}{p_i} + (1-\lambda_1)\overline{Q}_r\right)\right| \\
&\leq \frac{L}{P_r}\sum_{i=1}^{n} p_i r_i \left|\lambda_2 \frac{q_i}{p_i} + (1-\lambda_2)\overline{Q}_r - \lambda_1 \frac{q_i}{p_i} - (1-\lambda_1)\overline{Q}_r\right| \\
&= \frac{L|\lambda_2 - \lambda_1|}{P_r}\sum_{i=1}^{n} p_i r_i \left|\frac{q_i}{p_i} - \overline{Q}_r\right|
\end{aligned}$$

and get $|H(\lambda_2) - H(\lambda_1)| \leq L_1|\lambda_2 - \lambda_1|$, for L_1 defined by (10.47). For $\lambda_1 = 1$ and $\lambda_2 = \lambda$, left hand side in (10.48) is equal to $|H(\lambda) - H(1)|$. Since we already proved H is L_1-Lipschitzian function, inequality (10.48) holds. Analogously, (10.49) follows for $\lambda_1 = \lambda$ and $\lambda_2 = 0$. Finally, inequality (10.50) follows from (10.48) and (10.49),

$$\begin{aligned}
&\left|H(\lambda) - \frac{\lambda}{P_r}\sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right) - (1-\lambda) f(\overline{Q}_r)\right| \leq \\
&\leq \left|\lambda H(\lambda) - \frac{\lambda}{P_r}\sum_{i=1}^{n} r_i p_i f\left(\frac{q_i}{p_i}\right)\right| + \left|-(1-\lambda) f\left(\frac{1}{P_r}\sum_{i=1}^{n} r_i q_i\right) + (1-\lambda) H(\lambda)\right| \\
&\leq 2\lambda(1-\lambda) L_1.
\end{aligned}$$

□

Theorem 10.6 *Let $\mathbf{p}, \boldsymbol{r} \in \mathbb{R}_+^n$, $\boldsymbol{q} \in \mathbb{R}^n$ and $f\colon \mathbb{R} \to \mathbb{R}$ be a L-Lipschitzian function. For a mapping $F\colon [0,1] \to \mathbb{R}$ defined by*

$$F(\lambda) = \frac{1}{P_r^2}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i r_i p_j r_j f\left(\lambda \frac{q_i}{p_i} + (1-\lambda)\frac{q_j}{p_j}\right) \tag{10.51}$$

we have the following:

(i) the mapping F is symmetric, i.e. $F(\lambda) = F(1-\lambda), \lambda \in [0,1]$

(ii) the mapping F is L_2-Lipschitzian on $[0,1]$, where

$$L_2 = \frac{L}{P_r^2}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i r_i p_j r_j \left|\frac{q_i}{p_i} - \frac{q_j}{p_j}\right| \tag{10.52}$$

(iii) the inequalities

$$\left|F(\lambda) - \frac{1}{P_r^2}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i r_i p_j r_j f\left[\frac{1}{2}\left(\frac{q_i}{p_i} + \frac{q_j}{p_j}\right)\right]\right| \leq \frac{L_2}{2}|2\lambda - 1| \tag{10.53}$$

and

$$\left|F(\lambda)-\frac{1}{P_r}\sum_{i=1}^{n}p_ir_if\left(\frac{q_i}{p_i}\right)\right|\le L_2\min\{\lambda,1-\lambda\} \tag{10.54}$$

hold for all $\lambda\in[0,1]$.

Proof. The first property follows immediately from the definition (10.51). For proving the next property, let $\lambda_1,\lambda_2\in[0,1]$. Then we have

$$\begin{aligned}
&|F(\lambda_2)-F(\lambda_1)|\\
&=\frac{1}{P_r^2}\left|\sum_{i=1}^{n}\sum_{j=1}^{n}p_ir_ip_jr_j\left[f\left(\lambda_2\frac{q_i}{p_i}+(1-\lambda_2)\frac{q_j}{p_j}\right)-f\left(\lambda_1\frac{q_i}{p_i}+(1-\lambda_1)\frac{q_j}{p_j}\right)\right]\right|\\
&\le\frac{1}{P_r^2}\sum_{i=1}^{n}\sum_{j=1}^{n}p_iri_ip_jr_j\left|f\left(\lambda_2\frac{q_i}{p_i}+(1-\lambda_2)\frac{q_j}{p_j}\right)-f\left(\lambda_1\frac{q_i}{p_i}+(1-\lambda_1)\frac{q_j}{p_j}\right)\right|\\
&\le\frac{L|\lambda_2-\lambda_1|}{P_r^2}\sum_{i=1}^{n}\sum_{j=1}^{n}p_ir_ip_jr_j\left|\frac{q_i}{p_i}-\frac{q_j}{p_j}\right|\\
&=L_2|\lambda_2-\lambda_1|.
\end{aligned}$$

Inequality (10.53) follows from Lipschitzian property of F for $\lambda_1=\frac{1}{2}$ and $\lambda_2=\lambda$. So, we have $|F(\lambda)-F(\frac{1}{2})|\le L_2|\lambda-\frac{1}{2}|=\frac{L_2}{2}|\lambda-1|$. Analogously, (10.54) follows for $\lambda_1=1,\lambda_2=\lambda$ and $\lambda_1=1,\lambda_2=1-\lambda$. Namely, by combining

$$|F(\lambda)-F(1)|=\left|F(\lambda)-\frac{1}{P_r}\sum_{i=1}^{n}p_ir_if\left(\frac{q_i}{p_i}\right)\right|\le|\lambda-1|=1-\lambda$$

and

$$|F(1-\lambda)-F(1)|=|F(\lambda)-F(1)|\le|1-\lambda-1|=\lambda$$

we get (10.54). □

The next result offers us the relation between the mappings F and H, defined by (10.46) and (10.51).

Theorem 10.7 *For mappings* $F\colon[0,1]\to\mathbb{R}$ *and* $H\colon[0,1]\to\mathbb{R}$ *defined by* (10.51) *and* (10.46)*, inequality*

$$|F(\lambda)-H(\lambda)|\le(1-\lambda)L_1 \tag{10.55}$$

holds for all $\lambda\in[0,1]$*, where* L_1 *is defined by* (10.47).

Proof. The inequality (10.55) holds as follows:

$$\begin{aligned}
&|F(\lambda)-H(\lambda)|=\\
&=\left|\frac{1}{P_r^2}\sum_{i=1}^{n}\sum_{j=1}^{n}p_ir_ip_jr_jf\left(\lambda\frac{q_i}{p_i}+(1-\lambda)\frac{q_j}{p_j}\right)-\frac{1}{P_r}\sum_{i=1}^{n}p_ir_if\left(\lambda\frac{q_i}{p_i}+(1-\lambda)\overline{Q}_r\right)\right|
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{P_r^2}\left|\sum_{i=1}^{n}\sum_{j=1}^{n} p_i r_i p_j r_j \left[f\left(\lambda\frac{q_i}{p_i}+(1-\lambda)\frac{q_j}{p_j}\right)-f\left(\lambda\frac{q_i}{p_i}+(1-\lambda)\overline{Q}_r\right)\right]\right| \\
&\le \frac{1}{P_r^2}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i r_i p_j r_j \left|f\left(\lambda\frac{q_i}{p_i}+(1-\lambda)\frac{q_j}{p_j}\right)-f\left(\lambda\frac{q_i}{p_i}+(1-\lambda)\overline{Q}_r\right)\right| \\
&\le \frac{L}{P_r^2}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i r_i p_j r_j \left|\lambda\frac{q_i}{p_i}+(1-\lambda)\frac{q_j}{p_j}-\lambda\frac{q_i}{p_i}-(1-\lambda)\overline{Q}_r\right| \\
&= \frac{L(1-\lambda)}{P_r}\sum_{i=1}^{n} p_i r_i \left|\frac{q_i}{p_i}-\overline{Q}_r\right| = (1-\lambda)L_1.
\end{aligned}
$$

□

10.6 Applications to Zipf-Mandelbrot law

Definition 10.1 *[11] Zipf-Mandelbrot law is a discrete probability distribution, depends on three parameters $N \in \{1,2,\ldots\}$, $t \in [0,\infty\rangle$ and $v > 0$, and it is defined by*

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

where

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

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

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

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

$$p_i = \phi(i;N,t_1,v_1) := \frac{1}{(i+t_1)^{v_1} H_{N,t_1,v_1}}, \quad i = 1,\ldots,N, \tag{10.56}$$

and

$$q_i = \phi(i;N,t_2,v_2) := \frac{1}{(i+t_2)^{v_2} H_{N,t_2,v_2}}, \quad i = 1,\ldots,N, \tag{10.57}$$

where

$$H_{N,t_k,v_k} := \sum_{j=1}^{N} \frac{1}{(j+t_k)^{v_k}}, \quad k = 1,2. \tag{10.58}$$

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

$$C_f(\mathbf{q},\mathbf{p};\mathbf{r}) = \frac{1}{H_{N,t_1,v_1}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} f\left(\frac{(i+t_1)^{v_1} H_{N,t_1,v_1}}{(i+t_2)^{v_2} H_{N,t_2,v_2}}\right). \tag{10.59}$$

Using (10.56) and (10.57), we have the following expressions for (10.17) and (10.18)

$$P_r = \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1} H_{N,t_1,v_1}} = \frac{1}{H_{N,t_1,v_1}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}, \tag{10.60}$$

$$\overline{Q}_r = \frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2} H_{N,t_2,v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1} H_{N,t_1,v_1}}} = \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} \cdot \frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}}. \tag{10.61}$$

For m and M from Theorem 10.3 we have

$$m = \frac{(1+t_1)^{v_1} H_{N,t_1,v_1}}{(N+t_2)^{v_2} H_{N,t_2,v_2}}$$

and

$$M = \frac{(N+t_1)^{v_1} H_{N,t_1,v_1}}{(1+t_2)^{v_2} H_{N,t_2,v_2}}.$$

Thus we have the following results.

Corollary 10.1 *Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geq 0$ and $v_1, v_2 > 0$, respectively, and $\mathbf{r} \in \mathbb{R}^n_+$. If $C_f(\mathbf{q},\mathbf{p};\mathbf{r}), P_r$ and $\overline{Q}_r$ are defined by (10.59), (10.60) and (10.61), respectively, we have*

$$\left|C_f(\mathbf{q},\mathbf{p};\mathbf{r}) - P_r f(\overline{Q}_r)\right| \leq \frac{L}{H_{N,t_2,v_2}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} \left| \frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}} - \frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}} \right|,$$

and

$$\begin{aligned} &\left| \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} \left[\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}} - \frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}} \right) f\left(\frac{(1+t_1)^{v_1} H_{N,t_1,v_1}}{(N+t_2)^{v_2} H_{N,t_2,v_2}} \right) \right.\right. \\ &+ \left. \left(\frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}} - \frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}} \right) f\left(\frac{(N+t_1)^{v_1} H_{N,t_1,v_1}}{(1+t_2)^{v_2} H_{N,t_2,v_2}} \right) \right] \\ &- \left. \left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}} - \frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}} \right) C_f(\mathbf{q},\mathbf{p};\mathbf{r}) \right| \\ &\leq 2L \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} \left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}} - \frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}} \right) \left(\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}} - \frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}} \right). \end{aligned} \tag{10.62}$$

Proof. Inequality (10.62) can be obtained from

$$\left| \left[\left(\frac{(N+t_1)^{v_1} H_{N,t_1,v_1}}{(1+t_2)^{v_2} H_{N,t_2,v_2}} - \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} \frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}} \right) f\left(\frac{(1+t_1)^{v_1} H_{N,t_1,v_1}}{(N+t_2)^{v_2} H_{N,t_2,v_2}} \right) + \right.\right.$$

$$\left.\left(\frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}}\frac{\sum_{i=1}^{N}\frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}}-\frac{(1+t_1)^{v_1}H_{N,t_1,v_1}}{(N+t_2)^{v_2}H_{N,t_2,v_2}}\right)f\left(\frac{(N+t_1)^{v_1}H_{N,t_1,v_1}}{(1+t_2)^{v_2}H_{N,t_2,v_2}}\right)\right]\times$$

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

$$\leq 2L\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}H_{N,t_1,v_1}}\left(\frac{(N+t_1)^{v_1}H_{N,t_1,v_1}}{(1+t_2)^{v_2}H_{N,t_2,v_2}}-\frac{(i+t_1)^{v_1}H_{N,t_1,v_1}}{(i+t_2)^{v_2}H_{N,t_2,v_2}}\right)\times$$

$$\left(\frac{(i+t_1)^{v_1}H_{N,t_1,v_1}}{(i+t_2)^{v_2}H_{N,t_2,v_2}}-\frac{(1+t_1)^{v_1}H_{N,t_1,v_1}}{(N+t_2)^{v_2}H_{N,t_2,v_2}}\right).$$

□

Corollary 10.2 *Let the functions $G_k : [\alpha,\beta]\times[\alpha,\beta]\to\mathbb{R}$ $(k=0,1,2,3,4)$ be as defined in* (10.3)-(10.7)*. Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N\in\{1,2,\ldots\}$, t_1, $t_2\geq 0$ and $v_1,v_2>0$, respectively, and $\mathbf{r}\in\mathbb{R}_+^n$ such that*

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

and that $\overline{Q}_r\in[\alpha,\beta]$, where $\overline{Q}_r$ is as defined in (10.61).

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

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

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

holds, and

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

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

holds, where id is the identity function, Q, p_i, q_i, P_r, $C_f(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (10.22), (10.56), (10.57), (10.60), (10.59) respectively.

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

$$KL(\mathbf{q},\mathbf{p};\mathbf{r})=\frac{1}{H_{N,t_2,v_2}}\sum_{i=1}^{N}\frac{r_i}{(i+t_2)^{v_2}}\log\left(\frac{(i+t_1)^{v_1}H_{N,t_1,v_1}}{(i+t_2)^{v_2}H_{N,t_2,v_2}}\right). \tag{10.63}$$

The following results hold.

Corollary 10.3 *Let* $\mathbf{p},\mathbf{q}$ *be two Zipf-Mandelbrot laws with parameters* $N \in \{1,2,\ldots\}$, t_1, $t_2 \geq 0$ *and* $v_1, v_2 > 0$, *respectively, and* $\mathbf{r} \in \mathbb{R}^n_+$. *If* $KL(\mathbf{q},\mathbf{p};\mathbf{r})$ *is defined by (10.63), then inequalities*

$$\begin{aligned}
&\Bigg| H_{N,t_2,v_2} KL(\mathbf{q},\mathbf{p};\mathbf{r}) - \sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}} \Bigg(\log \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} + \log \sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}} - \\
&- \log \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} \Bigg) \Bigg| \leq \\
&\leq \max \left\{ \left| \log \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} + s_1 \log(1+t_1) - v_2 \log(N+t_2) + 1 \right|, \right. \\
&\qquad \left. \left| \log \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} + v_1 \log(N+t_1) - v_2 \log(1+t_2) + 1 \right| \right\} \times \\
&\qquad \times \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} \left| \frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}} - \frac{\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}}} \right|,
\end{aligned}$$

and

$$\begin{aligned}
&\Bigg| \frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}} \left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} - \sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}} \right) \\
&+ \frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}} \left(\sum_{i=1}^{N} \frac{r_i}{(i+t_2)^{v_2}} - \frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}} \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} \right) \\
&- H_{N,t_2,v_2} \left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}} - \frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}} \right) KL(\mathbf{q},\mathbf{p};\mathbf{r}) \Bigg| \\
&\leq 2 \max \left\{ \left| \log \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} + v_1 \log(1+t_1) - v_2 \log(N+t_2) + 1 \right|, \right. \\
&\qquad \left. \left| \log \frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}} + v_1 \log(N+t_1) - v_2 \log(1+t_2) + 1 \right| \right\} \times \\
&\qquad \times \sum_{i=1}^{N} \frac{r_i}{(i+t_1)^{v_1}} \left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}} - \frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}} \right) \left(\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}} - \frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}} \right)
\end{aligned}$$

hold.

Corollary 10.4 *Let the functions* $G_k : [\alpha,\beta] \times [\alpha,\beta] \to \mathbb{R}$ $(k = 0,1,2,3,4)$ *be as defined in* (10.3)-(10.7). *Let* $\mathbf{p},\mathbf{q}$ *be two Zipf-Mandelbrot laws with parameters* $N \in \{1,2,\ldots\}$, t_1, $t_2 \geq 0$ *and* $v_1, v_2 > 0$, *respectively, and* $\mathbf{r} \in \mathbb{R}^n_+$ *such that*

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

and that $\overline{Q}_r \in [\alpha,\beta]$, *where* $\overline{Q}_r$ *is as defined in (10.61).*

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

Then

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

holds, where id is the identity function, Q, p_i, q_i, P_r, $KL(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (10.22), (10.56), (10.57), (10.60), (10.63) respectively.

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

$$He(\mathbf{q},\mathbf{p};\mathbf{r})=\frac{1}{H_{N,t_1,v_1}H_{N,t_2,v_2}}\sum_{i=1}^N r_i\frac{\left(\sqrt{(i+t_1)^{v_1}H_{N,t_1,v_1}}-\sqrt{(i+t_2)^{v_2}H_{N,t_2,v_2}}\right)^2}{(i+t_1)^{v_1}(i+t_2)^{v_2}}. \tag{10.64}$$

The following results hold true.

Corollary 10.5 *Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N\in\{1,2,\ldots\}$, t_1, $t_2\geq 0$ and $v_1,v_2>0$, respectively, and $\mathbf{r}\in\mathbb{R}_+^n$. If $He(\mathbf{q},\mathbf{p};\mathbf{r})$ is defined by (10.64), then inequalities*

$$\begin{aligned}&\left|He(\mathbf{q},\mathbf{p};\mathbf{r})-\left(\sqrt{\sum_{i=1}^N\frac{r_i}{(i+t_1)^{v_1}H_{N,t_1,v_1}}}-\sqrt{\sum_{i=1}^N\frac{r_i}{(i+t_2)^{v_2}H_{N,t_2,v_2}}}\right)^2\right|\\&\leq\max\left\{\left|1-\sqrt{\frac{(N+t_2)^{v_2}H_{N,t_2,v_2}}{(1+t_1)^{v_1}H_{N,t_1,v_1}}}\right|,\left|1-\sqrt{\frac{(1+t_2)^{v_2}H_{N,t_2,v_2}}{(N+t_1)^{v_1}H_{N,t_1,v_1}}}\right|\right\}\times\\&\quad\times\sum_{i=1}^N\frac{r_i}{(i+t_1)^{v_1}H_{N,t_2,v_2}}\left|\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}}-\frac{\sum_{i=1}^N\frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^N\frac{r_i}{(i+t_1)^{v_1}}}\right|,\end{aligned}$$

and

$$\begin{aligned}&\left|\left[\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}}\sum_{i=1}^N\frac{r_i}{(i+t_1)^{v_1}}-\sum_{i=1}^N\frac{r_i}{(i+t_2)^{v_2}}\right)\left(1-\sqrt{\frac{(1+t_1)^{v_1}H_{N,t_1,v_1}}{(N+t_2)^{v_2}H_{N,t_2,v_2}}}\right)^2\right.\right.\\&+\left.\left(\sum_{i=1}^N\frac{r_i}{(i+t_2)^{v_2}}-\frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}}\sum_{i=1}^N\frac{r_i}{(i+t_1)^{v_1}}\right)\left(1-\sqrt{\frac{(N+t_1)^{v_1}H_{N,t_1,v_1}}{(1+t_2)^{v_2}H_{N,t_2,v_2}}}\right)^2\right]\\&-\left.H_{N,t_1,v_1}\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}}-\frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}}\right)He(\mathbf{q},\mathbf{p};\mathbf{r})\right|\\&\leq 2\frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}}\max\left\{\left|1-\sqrt{\frac{(N+t_2)^{v_2}H_{N,t_2,v_2}}{(1+t_1)^{v_1}H_{N,t_1,v_1}}}\right|,\left|1-\sqrt{\frac{(1+t_2)^{v_2}H_{N,t_2,v_2}}{(N+t_1)^{v_1}H_{N,t_1,v_1}}}\right|\right\}\times\\&\quad\times\sum_{i=1}^N\frac{r_i}{(i+t_1)^{v_1}}\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}}-\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}}\right)\left(\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}}-\frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}}\right)\end{aligned}$$

hold.

Corollary 10.6 *Let the functions $G_k : [\alpha,\beta]\times[\alpha,\beta]\to\mathbb{R}$ $(k=0,1,2,3,4)$ be as defined in* (10.3)-(10.7). *Let* $\mathbf{p},\mathbf{q}$ *be two Zipf-Mandelbrot laws with parameters $N\in\{1,2,\dots\}$, t_1, $t_2\geq 0$ and $v_1,v_2>0$, respectively, and $\mathbf{r}\in\mathbb{R}^n_+$ such that*

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

and that $\overline{Q}_r\in[\alpha,\beta]$, where $\overline{Q}_r$ is as defined in (10.61).

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

Then

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

holds, where Q, p_i, q_i, P_r, $He(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (10.22), (10.56), (10.57), (10.60), (10.64) respectively, and $f(t)=(1-\sqrt{t})^2$, $t>0$.

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

$$Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r})=\frac{H^{\gamma-1}_{N,t_1,v_1}}{H^{\gamma}_{N,t_2,v_2}}\sum_{i=1}^{N}r_i\frac{(i+t_1)^{(\gamma-1)v_1}}{(i+t_2)^{\gamma v_2}}. \tag{10.65}$$

The following results hold.

Corollary 10.7 *Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N\in\{1,2,\dots\}$, t_1, $t_2\geq 0$ and $v_1,v_2>0$, respectively, and $\mathbf{r}\in\mathbb{R}^n_+$. If $Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r})$ is defined by (10.65), then inequalities*

$$\begin{aligned}&\left|\frac{H^{\gamma}_{N,t_2,v_2}}{H^{\gamma-1}_{N,t_1,v_1}}Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r})-\left(\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}\right)^{1-\gamma}\left(\sum_{i=1}^{N}\frac{r_i}{(i+t_2)^{v_2}}\right)^{\gamma}\right|\\&\leq\gamma\frac{(N+t_1)^{(\gamma-1)v_1}}{(1+t_2)^{(\gamma-1)v_2}}\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}\left|\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}}-\frac{\sum_{i=1}^{N}\frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}}\right|,\end{aligned}$$

and

$$\begin{aligned}&\left|\frac{(1+t_1)^{\gamma v_1}}{(N+t_2)^{\gamma v_2}}\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}}\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}-\sum_{i=1}^{N}\frac{r_i}{(i+t_2)^{v_2}}\right)\right.\\&+\frac{(N+t_1)^{\gamma v_1}}{(1+t_2)^{\gamma v_2}}\left(\sum_{i=1}^{N}\frac{r_i}{(i+t_2)^{v_2}}-\frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}}\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}\right)\\&\left.-\frac{H^{\gamma}_{N,t_2,v_2}}{H^{\gamma-1}_{N,t_1,v_1}}\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}}-\frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}}\right)Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r})\right|\\&\leq 2\gamma\frac{(N+t_1)^{(\gamma-1)v_1}}{(1+t_2)^{(\gamma-1)v_2}}\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}}-\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}}\right)\times\end{aligned}$$

$$\times \left(\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}} - \frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}} \right)$$

hold.

Corollary 10.8 *Let the functions $G_k : [\alpha,\beta] \times [\alpha,\beta] \to \mathbb{R}$ $(k = 0,1,2,3,4)$ be as defined in* (10.3)-(10.7). *Let* $\mathbf{p},\mathbf{q}$ *be two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geq 0$ and $v_1, v_2 > 0$, respectively, and* $\mathbf{r} \in \mathbb{R}^n_+$ *such that*

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

and that $\overline{Q}_r \in [\alpha,\beta]$, where $\overline{Q}_r$ is as defined in (10.61).

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

Then

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

holds, where Q, p_i, q_i, P_r, $Re_\gamma(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (10.22), (10.56), (10.57), (10.60), (10.65) respectively, and $f(t) = t^\gamma$, $(t > 0, \gamma > 1)$.

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

$$D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}) = H_{N,t_1,v_1} \cdot \sum_{i=1}^N r_i (i+t_1)^{v_1} \left(\frac{1}{(i+t_2)^{v_2} H_{N,t_2,v_2}} - \frac{1}{(i+t_1)^{v_1} H_{N,t_1,v_1}} \right)^2. \quad (10.66)$$

We have the following results.

Corollary 10.9 *Let* $\mathbf{p},\mathbf{q}$ *be two Zipf-Mandelbrot laws with parameters $N \in \{1,2,\ldots\}$, t_1, $t_2 \geq 0$ and $v_1, v_2 > 0$, respectively, and* $\mathbf{r} \in \mathbb{R}^n_+$. *If $D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r})$ is defined by* (10.66), *then inequalities*

$$\begin{aligned}
&\left| D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r}) - \frac{H_{N,t_1,v_1}}{\sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}}} \left(\sum_{i=1}^N \frac{r_i}{(i+t_2)^{v_2} H_{N,t_2,v_2}} - \sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1} H_{N,t_1,v_1}} \right)^2 \right| \\
&\leq \frac{2}{H_{N,t_2,v_2}} \max\left\{ \left| \frac{(1+t_1)^{v_1} H_{N,t_1,v_1}}{(N+t_2)^{v_2} H_{N,t_2,v_2}} - 1 \right|, \left| \frac{(N+t_1)^{v_1} H_{N,t_1,v_1}}{(1+t_2)^{v_2} H_{N,t_2,v_2}} - 1 \right| \right\} \times \\
&\quad \times \sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}} \left| \frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}} - \frac{\sum_{i=1}^N \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}}} \right|
\end{aligned}$$

and

$$\left| \sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}} \left[\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}} - \frac{\sum_{i=1}^N \frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}}} \right) \left(\frac{(1+t_1)^{v_1} H_{N,t_1,v_1}}{(N+t_2)^{v_2} H_{N,t_2,v_2}} - 1 \right)^2 \right. \right.$$

$$
\begin{aligned}
&+\left(\frac{\sum_{i=1}^{N}\frac{r_i}{(i+t_2)^{v_2}}}{\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}}-\frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}}\right)\left(\frac{(N+t_1)^{v_1}H_{N,t_1,v_1}}{(1+t_2)^{v_2}H_{N,t_2,v_2}}-1\right)^2\Bigg]\\
&-\left.D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r})H_{N,t_1,v_1}\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}}-\frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}}\right)\right|\\
&\leq 4\left(\frac{H_{N,t_1,v_1}}{H_{N,t_2,v_2}}\right)^2\max\left\{\left|\frac{(1+t_1)^{v_1}H_{N,t_1,v_1}}{(N+t_2)^{v_2}H_{N,t_2,v_2}}-1\right|,\left|\frac{(N+t_1)^{v_1}H_{N,t_1,v_1}}{(1+t_2)^{v_2}H_{N,t_2,v_2}}-1\right|\right\}\times\\
&\quad\times\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}\left(\frac{(N+t_1)^{v_1}}{(1+t_2)^{v_2}}-\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}}\right)\left(\frac{(i+t_1)^{v_1}}{(i+t_2)^{v_2}}-\frac{(1+t_1)^{v_1}}{(N+t_2)^{v_2}}\right)
\end{aligned}
$$

hold.

Corollary 10.10 *Let the functions $G_k:[\alpha,\beta]\times[\alpha,\beta]\to\mathbb{R}$ $(k=0,1,2,3,4)$ be as defined in* (10.3)-(10.7). *Let $\mathbf{p},\mathbf{q}$ be two Zipf-Mandelbrot laws with parameters $N\in\{1,2,\ldots\}$, t_1, $t_2\geq 0$ and $v_1,v_2>0$, respectively, and $\mathbf{r}\in\mathbb{R}_+^n$ such that*

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

and that $\overline{Q}_r\in[\alpha,\beta]$, where $\overline{Q}_r$ is as defined in (10.61).

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

Then

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

holds, where Q, p_i, q_i, P_r, $D_{\chi^2}(\mathbf{q},\mathbf{p};\mathbf{r})$ are as defined in (10.22), (10.56), (10.57), (10.60), (10.66) respectively, and $f(t)=(t-1)^2,(t>0)$.

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

$$H(\mathbf{p};\mathbf{r})=\frac{1}{H_{N,t_1,v_1}}\sum_{i=1}^{N}\frac{r_i}{(i+t_1)^{v_1}}\log\left[(i+t_1)^{v_1}H_{N,t_1,v_1}\right]. \tag{10.67}$$

We have the following results.

Corollary 10.11 *Let $\mathbf{p}$ be the Zipf-Mandelbrot law with parameters $N\in\{1,2,\ldots\}$, $t_1\geq 0$ and $v_1>0$, and $\mathbf{r}\in\mathbb{R}_+^n$. If $H(\mathbf{p};\mathbf{r})$ is defined by (10.67), then inequalities*

$$
\begin{aligned}
&\left|H_{N,t_1,v_1}H(\mathbf{p};\mathbf{r})-\sum_{i=1}^{N}\frac{r_i}{(1+t_1)^{v_1}}\log\left(H_{N,t_1,v_1}\frac{\sum_{i=1}^{N}r_i}{\sum_{i=1}^{N}\frac{r_i}{(1+t_1)^{v_1}}}\right)\right|\\
&\leq\frac{1}{(1+t_1)^{v_1}}\sum_{i=1}^{N}\frac{r_i}{(1+t_1)^{v_1}}\left|(1+t_1)^{v_1}-\frac{\sum_{i=1}^{N}r_i}{\sum_{i=1}^{N}\frac{r_i}{(1+t_1)^{v_1}}}\right|
\end{aligned}
$$

and

$$\left| \left[(N+t_1)^{v_1} - \frac{\sum_{i=1}^N r_i}{\sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}}} \right] f\left((1+t_1)^{v_1} H_{N,t_1,v_1}\right) + \left[\frac{\sum_{i=1}^n r_i}{\sum_{i=1}^n \frac{r_i}{(i+t_1)^{v_1}}} - (1+t_1)^{v_1} \right] \times \right.$$
$$\left. \times\ f\left((N+t_1)^{v_1} H_{N,t_1,v_1}\right) - \frac{H_{N,t_1,v_1}}{\sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}}} \left[(N+t_1)^{v_1} - (1+t_1)^{v_1} \right] H(\mathbf{p};\mathbf{r}) \right|$$
$$\leq \frac{2}{(1+t_1)^{v_1} \sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}}} \times$$
$$\times \sum_{i=1}^N \frac{r_i}{(i+t_1)^{v_1}} \left[(N+t_1)^{v_1} - (i+t_1)^{v_1} \right] \left[(i+t_1)^{v_1} - (1+t_1)^{v_1} \right]$$

hold.

Corollary 10.12 *Let* $\mathbf{p}$ *be the Zipf-Mandelbrot law with parameters* $N \in \{1,2,\ldots\}$, $t_1 \geq 0$ *and* $v_1 > 0$, *and* $\mathbf{r} \in \mathbb{R}_+^n$ *such that*

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

$$\text{and that } \frac{1}{P_r} \sum_{i=1}^n r_i \in [\alpha,\beta], \text{ where } P_r \text{ is as defined in (10.60).}$$

Let the functions $G_k : [\alpha,\beta] \times [\alpha,\beta] \to \mathbb{R}$ $(k = 0,1,2,3,4)$ *be as defined in* (10.3)-(10.7). *Furthermore, let* $p,q \in \mathbb{R}$, $1 \leq p,q \leq \infty$, *be such that* $\frac{1}{p} + \frac{1}{q} = 1$.

Then

$$\left| \frac{1}{P_r} H(\mathbf{p};\mathbf{r}) - \log\left(\frac{1}{P_r} \sum_{i=1}^n r_i \right) \right| \leq Q \cdot \left\| \log'' \right\|_p \tag{10.68}$$

holds, where Q, p_i, q_i, $H(\mathbf{p};\mathbf{r})$ *are as defined in (10.22), (10.56), (10.57), (10.67) respectively.*

Bibliography

[1] S. M. Ali and S. D. Silvey, *A general class of coefficients of divergence of one distribution from another*, J. Roy. Statist. Soc. series B, **28** (1), (1966), 131–142.

[2] R. Beran, *Minimum Hellinger distance estimates for parametric models*, Ann. Statist., **5** (1977), 445–463.

[3] Y. J. Cho, M. Matić and J. Pečarić, *Inequalities of Jensens type for Lipschitzian mappings*, Comm. on Appl. Nonlinear Analysis **8** (2001), 37–46.

[4] Y. J. Cho, M. Matić and J. Pečarić, *Two mappings in connection to Jensens inequality*, Panamerican Math.J. **12**(1), (2002), 43–50.

[5] I. Csiszár, *Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Bewis der Ergodizität von Markhoffschen Ketten*, Publ. Math. Inst. Hungar. Acad. Sci. **8**, (1963), 85–108.

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

[7] S. S. Dragomir, *A mapping in connection to Hadamard's inequalities*, J. Math. Anal. Appl. **167** (1992), 49–56.

[8] S. S. Dragomir, *On Hadamard's inequalities for convex functions*, Mat. Balkanicae **6** (4), (1992), 215–222.

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

[10] S. S. Dragomir, Y. J. Cho and S. S. Kim, *Inequalities of Hadamand's type for Lipschitzian mappings and their applications*, J. Math. Anal. Appl. **245** (2000), 489–501.

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

[12] 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.

[13] S. Kullback, *Information Theory and Statistics*, J. Wiley, New York, 1959.

[14] S. Kullback and R. A. Leibler, *On information and sufficiency*, Annals Math. Statist., **12** (1951), 79–86.

[15] M. Matić, C. E. M. Pearce and J. Pečarić, *Shannon's and related inequalities in information theory*, Survey on classical inequalities, editor Themistocles M. Rassias, Kluwer Academic Publ., 2000, 127–164.

[16] M. Matić, C.E.M. Pearce and J.Pečarić, *Some comparison theorems for the mean-value characterization of "useful" information measures*, Southeast Asian Bulletin of Mathematics, **23** (1999), 111–116.

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

[18] T. Morimoto, *Markov processes and the H-theorem*, J. Phys. Soc. Jap. **18** (3), (1963), 328–331.

[19] Đ. Pečarić, J. Pečarić and D. Pokaz, *Generalized Csiszár's $f-$divergence for Lipschitzian functions*, Math. Inequal. Appl., submitted for publication.

[20] Đ. Pečarić, J. Pečarić and M. Rodić, *About the sharpness of the Jensen inequality*, J. Inequal. Appl. (2018) 2018:337, doi: 10.1186/s13660-018-1923-4.

[21] Đ. Pečarić, J. Pečarić and M. Rodić, *On a Jensen-type inequality for generalized f-divergences and Zipf-Mandelbrot law*, Math. Inequal. Appl., accepted.

[22] J. Pečarić, I. Perić and M. Rodić Lipanović, *Uniform treatment of Jensen type inequalities*, Math. Reports, **16** (66), 2 (2014), 183–205.

[23] J. Pečarić and M. Rodić, *Uniform treatment of Jensen type inequalities II.*, Math. Reports, accepted.

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

Chapter 11

On Shannon and Zipf-Mandelbrot entropies and related results

Sadia Khalid, Đilda Pečarić and Josip Pečarić

Abstract. We present some interesting results related to the bounds of Shannon and Zipf-Mandelbrot entropies. Further, we define linear functionals as the non-negative differences of the obtained inequalities and present mean value theorems. We also discuss the properties of the functionals, such as n-exponential and logarithmic convexity. Finally, we present examples of the family of functions for which the results can be applied.

11.1 Introduction and Preliminaries

Definition 11.1 *The Shannon entropy of a positive probability distribution* $\mathbf{p} = (p_1, \dots, p_n)$ *is defined by* $S(\mathbf{p}) := \sum_{k=1}^{n} p_k \log\left(\frac{1}{p_k}\right)$.

An important result related to the bounds of the Shannon entropy is given in [4].

Theorem 11.1 *Let $p_k > 0$ $(1 \le k \le n)$ be a probability distribution with Shannon entropy $S(\mathbf{p})$ and $P_k = \sum_{i=1}^{k} p_i$ $(1 \le k \le n)$. Then*

$$\begin{aligned}
&S(\mathbf{p}) + \sum_{k=2}^{n} ((k-1)p_k - P_{k-1})(\log k - \log(k-1)) \\
&\le \sum_{k=2}^{n} F(k-1)\, p_k \\
&\le S(\mathbf{p}) + \sum_{k=2}^{n} ((k-1)p_k - P_{k-1})(\log P_k - \log P_{k-1}), \qquad (11.1)
\end{aligned}$$

where

$$F(x) = (x+1)\log(x+1) - x\log x \, (x>0). \qquad (11.2)$$

Equalities hold in (11.1) *if $p_k = \frac{1}{n}$ $(1 \le k \le n)$.*

S. Khalid, J. Pečarić and M. Praljak presented the following generalization of Throrem 11.1 in [6, Theorems 2.1 and 2.3].

Theorem 11.2 *Let $a_k > 0$ and $p_k > 0$ $(1 \le k \le n)$ be real numbers such that $P_k = \sum_{i=1}^{k} p_i$ $(1 \le k \le n)$. Let $g : [a,b] \to \mathbb{R}$ be a differentiable function such that $g(x+h) - g(x)$ is convex for all $x, x+h \in [a,b]$, where $h \ge 0$.*

(i) If P_{k-1}, P_k, $\frac{\sum_{i=1}^{k-1} p_i a_i}{a_k}$, $\frac{\sum_{i=1}^{k} p_i a_i}{a_k} \in [a,b]$ for all $k \in \{2,\dots,n\}$, then for any $s \in \mathbb{R}$, we have

$$\begin{aligned}
&\sum_{k=2}^{n} a_k^s (g(P_k) - g(P_{k-1})) + \sum_{k=2}^{n} a_k^{s-1} \left(\sum_{i=1}^{k-1} p_i a_i - P_{k-1} a_k\right) \left(g'(P_k) - g'(P_{k-1})\right) \\
&\le \sum_{k=2}^{n} a_k^s \left(g\left(\frac{\sum_{i=1}^{k} p_i a_i}{a_k}\right) - g\left(\frac{\sum_{i=1}^{k-1} p_i a_i}{a_k}\right)\right) \\
&\le \sum_{k=2}^{n} a_k^s (g(P_k) - g(P_{k-1})) \\
&+ \sum_{k=2}^{n} a_k^{s-1} \left(\sum_{i=1}^{k-1} p_i a_i - P_{k-1} a_k\right) \left(g'\left(\frac{\sum_{i=1}^{k} p_i a_i}{a_k}\right) - g'\left(\frac{\sum_{i=1}^{k-1} p_i a_i}{a_k}\right)\right). \qquad (11.3)
\end{aligned}$$

(ii) If $p_1 a_1$, $P_{k-1} a_k$, $P_k a_k$, $\sum_{i=1}^{k-1} p_i a_i$, $\sum_{i=1}^{k} p_i a_i \in [a,b]$ for all $k \in \{2,\dots,n\}$, then we have

$$\begin{aligned}
&g(p_1 a_1) + \sum_{k=2}^{n} (g(P_k a_k) - g(P_{k-1} a_k)) \\
&+ \sum_{k=2}^{n} \left(\sum_{i=1}^{k-1} p_i a_i - P_{k-1} a_k\right) \left(g'(P_k a_k) - g'(P_{k-1} a_k)\right) \\
&\le g\left(\sum_{i=1}^{n} p_i a_i\right)
\end{aligned}$$

$$\leq g(p_1a_1) + \sum_{k=2}^{n} \left(g(P_k a_k) - g(P_{k-1}a_k)\right)$$
$$+ \sum_{k=2}^{n} \left(\sum_{i=1}^{k-1} p_i a_i - P_{k-1}a_k \right) \left(g' \left(\sum_{i=1}^{k} p_i a_i \right) - g' \left(\sum_{i=1}^{k-1} p_i a_i \right) \right). \qquad (11.4)$$

If $g(x+h) - g(x)$ is concave for all $x, x+h \in [a,b]$ such that $h \geq 0$, then the reversed inequalities hold in (11.3) *and* (11.4).

Divided difference of a function is defined as follows (see [15, p. 14]):

Definition 11.2 *The nth-order divided difference of a function $f : [a,b] \to \mathbb{R}$ at mutually distinct points $x_0, \ldots, x_n \in [a,b]$ is defined recursively by*

$$[x_i; f] = f(x_i), \qquad i \in \{0, \ldots, n\},$$
$$[x_0, \ldots, x_n; f] = \frac{[x_1, \ldots, x_n; f] - [x_0, \ldots, x_{n-1}; f]}{x_n - x_0}.$$

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

n-convex functions can be characterized by the nth-order divided difference (see [15, p. 15]).

Definition 11.3 *A function $f : [a,b] \to \mathbb{R}$ is said to be n-convex $(n \geq 0)$ if and only if for all choices of $(n+1)$ distinct points $x_0, \ldots, x_n \in [a,b]$, the nth-order divided difference of f satisfies $[x_0, \ldots, x_n; f] \geq 0$.*

Remark 11.1 Note that 0-convex functions are non-negative functions, 1-convex functions are increasing functions and 2-convex functions are simply the convex functions.

The following interesting result related to the 3-convexity of the function g, is also presented in [6, Corollaries 2.2 and 2.4].

Corollary 11.1 *Let $a_k > 0$ and $p_k > 0$ $(1 \leq k \leq n)$ be real numbers such that $P_k = \sum_{i=1}^{k} p_i$ $(1 \leq k \leq n)$ and let $g : [a,b] \to \mathbb{R}$ be a differentiable function.*

(i) Let P_{k-1}, P_k, $\frac{\sum_{i=1}^{k-1} p_i a_i}{a_k}$, $\frac{\sum_{i=1}^{k} p_i a_i}{a_k} \in [a,b]$ for all $k \in \{2, \ldots, n\}$. If g is 3-convex, then (11.3) *holds for any $s \in \mathbb{R}$.*

(ii) Let $p_1 a_1$, $P_{k-1}a_k$, $P_k a_k$, $\sum_{i=1}^{k-1} p_i a_i$, $\sum_{i=1}^{k} p_i a_i \in [a,b]$ for all $k \in \{2, \ldots, n\}$. If g is 3-convex, then (11.4) *holds.*

If g is 3-concave, then the reversed inequalities hold in (11.3) *and* (11.4).

A sequence $\{a_k\}_{k \in \mathbb{N}}$ of real numbers which is non-increasing in weighted mean (see [6]) can be defined as follows:

Definition 11.4 *A sequence $\{a_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ is non-increasing in weighted mean if*

$$\frac{1}{P_n}\sum_{k=1}^{n} p_k a_k \geq \frac{1}{P_{n+1}}\sum_{k=1}^{n+1} p_k a_k, \quad n \in \mathbb{N}, \tag{11.5}$$

where a_k and p_k $(k \in \mathbb{N})$ are real numbers such that $p_i > 0$ $(1 \leq i \leq k)$ with $P_k := \sum_{i=1}^{k} p_i$ $(k \in \mathbb{N})$.

If the reversed inequality holds in (11.5), then the sequence $\{a_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ is called non-decreasing in weighted mean.
In a similar way, we can define when a finite sequence $\{a_k\}_{k=1}^{n} \subset \mathbb{R}$ is non-increasing or non-decreasing in weighted mean.

In [1] G. Bennett proved the weighted version of an inequality presented by Hardy-Littlewood-Pólya (see [2, Theorem 134]) for power functions $f(x) = x^s$: if a_k $(1 \leq k \leq n)$ are non-negative and non-increasing and $p_k \geq 0$ for all $k \in \{1,\ldots,n\}$ such that $P_k = \sum_{i=1}^{k} p_i$ $(1 \leq k \leq n)$, then for any real number $s > 1$, the inequality

$$\left(\sum_{k=1}^{n} p_k a_k\right)^s \geq \sum_{k=1}^{n} P_k^s\left(a_k^s - a_{k+1}^s\right) = (p_1 a_1)^s + \sum_{k=2}^{n} a_k^s\left(P_k^s - P_{k-1}^s\right) \tag{11.6}$$

holds. If $0 < s < 1$, then the reversed inequality holds in (11.6) (see [1]).

S. Khalid, J. Pečarić and M. Praljak presented the following generalization of inequality (11.6) in [6, Theorem 3].

Theorem 11.3 *Let a_k and p_k $(1 \leq k \leq n)$ be real numbers such that $a_k \geq 0$ and $p_k > 0$. Let $p_1 a_1$, $\sum_{k=1}^{n} p_k a_k$, $P_k a_k$, $P_{k-1} a_k \in [a,b]$ for all $k \in \{2,\ldots,n\}$ and let $f : [a,b] \to \mathbb{R}$ be a Wright-convex function.*

(i) If the sequence $\{a_k\}_{k=1}^{n}$ is non-increasing in weighted mean, then

$$f\left(\sum_{k=1}^{n} p_k a_k\right) \geq f(p_1 a_1) + \sum_{k=2}^{n}\left[f(P_k a_k) - f(P_{k-1} a_k)\right]. \tag{11.7}$$

(ii) If the sequence $\{a_k\}_{k=1}^{n}$ is non-decreasing in weighted mean, then

$$f\left(\sum_{k=1}^{n} p_k a_k\right) \leq f(p_1 a_1) + \sum_{k=2}^{n}\left[f(P_k a_k) - f(P_{k-1} a_k)\right]. \tag{11.8}$$

If f is Wright-concave, then the reversed inequalities hold in (11.7) and (11.8).

Definition 11.5 *Zipf-Mandelbrot law is a discrete probability distribution depending on three parameters $n \in \mathbb{N}$, $r \geq 0$ and $t > 0$, and is defined as*

$$f(i;n,r,t) = \frac{1}{(i+r)^t H_{n,r,t}}, \qquad i \in \{1,\ldots,n\},$$

where f is known as the probability mass function and

$$H_{n,r,t} := \sum_{k=1}^{n} \frac{1}{(k+r)^t} \tag{11.9}$$

is the generalized harmonic number.

If we take $p_k = \frac{1}{(k+r)^t H_{n,r,t}}$ $(1 \leq k \leq n,\ r \geq 0, t > 0$ and $H_{n,r,t}$ is the same as defined in Definition 11.5) in $S(\mathbf{p})$, then simple calculations reveal that

$$\begin{aligned} \sum_{k=1}^{n} \frac{1}{(k+r)^t H_{n,r,t}} \log\left((k+r)^t H_{n,r,t}\right) &= \frac{t}{H_{n,r,t}} \sum_{k=1}^{n} \frac{\log(k+r)}{(k+r)^t} + \log(H_{n,r,t}) \\ &:= Z(r,t,H_{n,r,t}), \end{aligned}$$

where $Z(r,t,H_{n,r,t})$ is known as Zipf-Mandelbrot entropy.

The results related to Shannon entropy and Zipf-Mandelbrot law are topic of great interest see for example [3], [10] and [11]. Zipf-Mandelbrot law is revisited in the context of linguistics in [13] (see also [12]).

In the first section of this chapter we present some interesting results related to the bounds of Zipf-Mandelbrot entropy and the 3-convexity of the function. In the second section, we present some interesting results related to the bounds of Shannon entropy by using non-increasing (non-decreasing) sequence of real numbers and by applying Theorem 11.3. Further, we also present some results related to the bounds of Zipf-Mandelbrot entropy. In both the sections, we define linear functionals as the non-negative differences of the obtained inequalities and we present mean value theorems for the linear functionals. We also discuss the n-exponential convexity and the log-convexity of the functions associated with the linear functionals.

11.2 On Zipf-Mandelbrot entropy and 3-convex functions

The results presented in this section are given in [7]. This section is organized as follows: in Section 11.2.1, we present some interesting results related to Zipf-Mandelbrot entropy. In Section 11.2.2, we define linear functionals as the non-negative differences of the obtained inequalities and present mean value theorems for the linear functionals. In Section 11.2.3, we present the properties of functionals, such as n-exponential and logarithmic convexity. Finally, we give an example of the family of functions for which the results can be applied.

11.2.1 Inequalities related to Zipf-Mandelbrot entropy

The aim of this section is to present some interesting results related to Zipf-Mandelbrot entropy.
Now first we define the cumulative distribution function as follows:

$$C_{k,n,r,t} := \frac{H_{k,r,t}}{H_{n,r,t}}, \tag{11.10}$$

where $k \in \{1,\dots,n\}, n \in \mathbb{N}, r \geq 0,\ t > 0$ and $H_{n,r,t}$ is the same as defined in Definition 11.5. The first result of this section states that:

Theorem 11.4 *Let $Z(r,t,H_{n,r,t})$ be the Zipf-Mandelbrot entropy and F and $C_{k,n,r,t}$ be the same as defined in* (11.2) *and* (11.10) *respectively. Then*

$$\begin{aligned}
&Z(r,t,H_{n,r,t}) + \sum_{k=2}^{n}\left(\frac{k-1}{(k+r)^t H_{n,r,t}} - C_{k-1,n,r,t}\right)\log\left(\frac{k}{k-1}\right)\\
&\leq \frac{1}{H_{n,r,t}}\sum_{k=2}^{n}\frac{F(k-1)}{(k+r)^t}\\
&\leq Z(r,t,H_{n,r,t}) + \sum_{k=2}^{n}\left(\frac{k-1}{(k+r)^t H_{n,r,t}} - C_{k-1,n,r,t}\right)\log\left(\frac{H_{k,r,t}}{H_{k-1,r,t}}\right).
\end{aligned}$$

Proof. Take $p_k = \frac{1}{(k+r)^t H_{n,r,t}}$ $(1 \leq k \leq n,\ r \geq 0,\ t > 0)$ in (11.1), the result is immediate. □

The second main result of this section states that:

Theorem 11.5 *Let $a_k > 0$ be real numbers and $H_{n,r,t}$ and $C_{k,n,r,t}$ be the same as defined in* (11.9) *and* (11.10) *respectively. Let $g : [a,b] \to \mathbb{R}$ be a differentiable function such that $g(x+h) - g(x)$ is convex for all $x, x+h \in [a,b]$, where $h \geq 0$.*

(*i*) *Let $C_{k-1,n,r,t}, C_{k,n,r,t}, \frac{1}{a_k H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t}, \frac{1}{a_k H_{n,r,t}}\sum_{i=1}^{k}\frac{a_i}{(i+r)^t} \in [a,b]$ for all $k \in \{2,\dots,n\}$. Then for any $s \in \mathbb{R}$, we have*

$$\begin{aligned}
&\sum_{k=2}^{n} a_k^s\left(g(C_{k,n,r,t}) - g(C_{k-1,n,r,t})\right)\\
&+\sum_{k=2}^{n} a_k^{s-1}\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t} - a_k C_{k-1,n,r,t}\right)\left(g'(C_{k,n,r,t}) - g'(C_{k-1,n,r,t})\right)\\
&\leq \sum_{k=2}^{n} a_k^s\left(g\left(\frac{1}{a_k H_{n,r,t}}\sum_{i=1}^{k}\frac{a_i}{(i+r)^t}\right) - g\left(\frac{1}{a_k H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t}\right)\right)\\
&\leq \sum_{k=2}^{n} a_k^s\left(g(C_{k,n,r,t}) - g(C_{k-1,n,r,t})\right)\\
&+\sum_{k=2}^{n} a_k^{s-1}\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t} - a_k C_{k-1,n,r,t}\right)
\end{aligned}$$

$$\times\left(g'\left(\frac{1}{a_kH_{n,r,t}}\sum_{i=1}^{k}\frac{a_i}{(i+r)^t}\right)-g'\left(\frac{1}{a_kH_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t}\right)\right). \tag{11.11}$$

(*ii*) *Let* $\frac{a_1}{(1+r)^tH_{n,r,t}}$, $a_kC_{k-1,n,r,t}$, $a_kC_{k,n,r,t}$, $\frac{1}{H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t}$, $\frac{1}{H_{n,r,t}}\sum_{i=1}^{k}\frac{a_i}{(i+r)^t}\in[a,b]$ *for all* $k\in\{2,\ldots,n\}$. *Then*

$$\begin{aligned}
&g\left(\frac{a_1}{(1+r)^tH_{n,r,t}}\right)+\sum_{k=2}^{n}\left(g\left(a_kC_{k,n,r,t}\right)-g\left(a_kC_{k-1,n,r,t}\right)\right)\\
&+\sum_{k=2}^{n}\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t}-a_kC_{k-1,n,r,t}\right)\left(g'\left(a_kC_{k,n,r,t}\right)-g'\left(a_kC_{k-1,n,r,t}\right)\right)\\
&\le g\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{n}\frac{a_i}{(i+r)^t}\right)\\
&\le g\left(\frac{a_1}{(1+r)^tH_{n,r,t}}\right)+\sum_{k=2}^{n}\left(g\left(a_kC_{k,n,r,t}\right)-g\left(a_kC_{k-1,n,r,t}\right)\right)\\
&+\sum_{k=2}^{n}\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t}-a_kC_{k-1,n,r,t}\right)\\
&\times\left(g'\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{k}\frac{a_i}{(i+r)^t}\right)-g'\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t}\right)\right).
\end{aligned} \tag{11.12}$$

If $g(x+h)-g(x)$ *is concave for all* $x,x+h\in[a,b]$ *such that* $h\ge 0$, *then the reversed inequalities hold in* (11.11) *and* (11.12).

Proof.

(*i*) By taking $p_i=\frac{1}{(i+r)^tH_{n,r,t}}$ in (11.3), the result is immediate.

(*ii*) The idea of the proof is the same as discussed in (*i*).

□

Corollary 11.2 *Let* $a_k>0$ *be real numbers,* $H_{n,r,t}$ *and* $C_{k,n,r,t}$ *be the same as defined in* (11.9) *and* (11.10) *respectively and let* $g:[a,b]\to\mathbb{R}$ *be a differentiable function.*

(*i*) *Let the condition of Theorem* 11.5 (*i*) *holds. If g is* 3*-convex, then* (11.11) *holds for any* $s\in\mathbb{R}$.

(*ii*) *Let the condition of Theorem* 11.5 (*ii*) *holds. If g is* 3*-convex, then* (11.12) *holds.*

If g is 3*-concave, then the reversed inequalities hold in* (11.11) *and* (11.12).

11.2.2 Linear functionals and mean value theorems

Consider the inequalities (11.11) and (11.12) and define linear functionals Ψ_i ($i=$ 1-6) by the non-negative differences of the inequalities (11.11) and (11.12) as follows:

$$\begin{aligned}
\Psi_1(g) &= \sum_{k=2}^{n} a_k^s \left(g\left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) - g\left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right) \\
&\quad - \sum_{k=2}^{n} a_k^s \left(g\left(C_{k,n,r,t}\right) - g\left(C_{k-1,n,r,t}\right) \right) \\
&\quad - \sum_{k=2}^{n} a_k^{s-1} D_k \left(g'\left(C_{k,n,r,t}\right) - g'\left(C_{k-1,n,r,t}\right) \right),
\end{aligned} \tag{11.13}$$

$$\begin{aligned}
\Psi_2(g) &= \sum_{k=2}^{n} a_k^{s-1} \left(\frac{1}{H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} - a_k C_{k-1,n,r,t} \right) \\
&\quad \times \left(g'\left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) - g'\left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right) \\
&\quad - \sum_{k=2}^{n} a_k^{s-1} D_k \left(g'\left(C_{k,n,r,t}\right) - g'\left(C_{k-1,n,r,t}\right) \right),
\end{aligned} \tag{11.14}$$

$$\begin{aligned}
\Psi_3(g) &= \sum_{k=2}^{n} a_k^s \left(g\left(C_{k,n,r,t}\right) - g\left(C_{k-1,n,r,t}\right) \right) \\
&\quad - \sum_{k=2}^{n} a_k^s \left(g\left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) - g\left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right) \\
&\quad + \sum_{k=2}^{n} a_k^{s-1} D_k \left(g'\left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) - g'\left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right),
\end{aligned} \tag{11.15}$$

$$\begin{aligned}
\Psi_4(g) &= g\left(\frac{1}{H_{n,r,t}} \sum_{i=1}^{n} \frac{a_i}{(i+r)^t} \right) - g\left(\frac{a_1}{(1+r)^t H_{n,r,t}} \right) \\
&\quad - \sum_{k=2}^{n} \left(g\left(a_k C_{k,n,r,t}\right) - g\left(a_k C_{k-1,n,r,t}\right) \right) \\
&\quad - \sum_{k=2}^{n} D_k \left(g'\left(a_k C_{k,n,r,t}\right) - g'\left(a_k C_{k-1,n,r,t}\right) \right),
\end{aligned} \tag{11.16}$$

$$\begin{aligned}
\Psi_5(g) &= \sum_{k=2}^{n} \left(\frac{1}{H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} - a_k C_{k-1,n,r,t} \right) \\
&\quad \times \left(g'\left(\frac{1}{H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) - g'\left(\frac{1}{H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right) \\
&\quad - \sum_{k=2}^{n} D_k \left(g'\left(a_k C_{k,n,r,t}\right) - g'\left(a_k C_{k-1,n,r,t}\right) \right)
\end{aligned} \tag{11.17}$$

and

$$\begin{aligned}\Psi_6(g) &= g\left(\frac{a_1}{(1+r)^t H_{n,r,t}}\right) - g\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{n}\frac{a_i}{(i+r)^t}\right)\\ &+\sum_{k=2}^{n}\left(g\left(a_k C_{k,n,r,t}\right) - g\left(a_k C_{k-1,n,r,t}\right)\right)\\ &+\sum_{k=2}^{n} D_k\left(g'\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{k}\frac{a_i}{(i+r)^t}\right) - g'\left(\frac{1}{H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t}\right)\right), \quad (11.18)\end{aligned}$$

where $D_k = \frac{1}{H_{n,r,t}}\sum_{i=1}^{k-1}\frac{a_i}{(i+r)^t} - a_k C_{k-1,n,r,t}$.
If $g:[a,b]\to\mathbb{R}$ is differentiable and 3-convex, then Corollary 11.2 implies that

$$\Psi_i(g) \ge 0, \qquad i\in\{1,\ldots,6\}.$$

Now we give mean value theorems for the functionals $\Psi_i\,(i=1\text{-}6)$ as defined in (11.13)-(11.18). These theorems enable us to define various classes of means that can be expressed in terms of linear functionals.

Theorem 11.6 *Let $g:[a,b]\to\mathbb{R}$ be such that $g\in C^3([a,b])$ and let Ψ_i $(i=1\text{-}6)$ be linear functionals as defined in (11.13)-(11.18). Then there exists $\delta_i\in[a,b]$ such that*

$$\Psi_i(g) = \frac{g'''(\delta_i)}{6}\Psi_i(g_0), \qquad i\in\{1,\ldots,6\},$$

where $g_0(x) = x^3$.

Proof. The proof is analogous to the proof of Theorem 2.7 in [6]. □

The following theorem is a new analogue of the classical Cauchy mean value theorem, related to the functionals Ψ_i $(i=1\text{-}6)$ and it can be proven by following the proof of Theorem 2.8 in [6].

Theorem 11.7 *Let $g,h:[a,b]\to\mathbb{R}$ be such that $g,h\in C^3([a,b])$ and let Ψ_i $(i=1\text{-}6)$ be linear functionals as defined in (11.13)-(11.18). Then there exists $\delta_i\in[a,b]$ such that*

$$\frac{\Psi_i(g)}{\Psi_i(h)} = \frac{g'''(\delta_i)}{h'''(\delta_i)}, \qquad i\in\{1,\ldots,6\}, \quad (11.19)$$

provided that the denominators are non-zero.

Remark 11.2 (i) By taking $g(x) = x^s$ and $h(x) = x^q$ in (11.19), where $s,q\in\mathbb{R}\setminus\{0,1,2\}$ are such that $s\neq q$, we have

$$\delta_i^{s-q} = \frac{q(q-1)(q-2)\Psi_i(x^s)}{s(s-1)(s-2)\Psi_i(x^q)}, \qquad i\in\{1,\ldots,6\}.$$

(ii) If the inverse of the function g'''/h''' exists, then (11.19) implies that

$$\delta_i = \left(\frac{g'''}{h'''}\right)^{-1}\left(\frac{\Psi_i(g)}{\Psi_i(h)}\right), \qquad i\in\{1,\ldots,6\}.$$

11.2.3 n-exponential convexity and log-convexity

In this section first we will present some important definitions which will be useful further. In the sequel, let I be an open interval in $\mathbb{R}$.

The next four definitions are given in [14].

Definition 11.6 *A function $f : I \to \mathbb{R}$ is n-exponentially convex in the Jensen sense if*

$$\sum_{i,j=1}^{n} \varsigma_i \varsigma_j f\left(\frac{x_i + x_j}{2}\right) \geq 0$$

holds for every $\varsigma_i \in \mathbb{R}$ and $x_i \in I$ $(1 \leq i \leq n)$.

Definition 11.7 *A function $f : I \to \mathbb{R}$ is n-exponentially convex if it is n-exponentially convex in the Jensen sense and continuous on I.*

Definition 11.8 *A function $f : I \to \mathbb{R}$ is exponentially convex in the Jensen sense if it is n-exponentially convex in the Jensen sense for all $n \in \mathbb{N}$.*

Definition 11.9 *A function $f : I \to \mathbb{R}$ is exponentially convex if it is exponentially convex in the Jensen sense and continuous.*

A log-convex function is defined as follows (see [15, p. 7]):

Definition 11.10 *A function $f : I \to (0,\infty)$ is said to be* log-convex *or multiplicatively convex if* $\log f$ *is convex. Equivalently, f is* log-convex *if for all $x, y \in I$ and for all $\lambda \in [0,1]$, the inequality*

$$f(\lambda x + (1-\lambda)y) \leq f^{\lambda}(x) f^{(1-\lambda)}(y)$$

holds. If the inequality reverses, then f is said to be log-concave.

Next we study the n-exponential convexity and log-convexity of the functions associated with the linear functionals Ψ_i $(i = 1\text{-}6)$ as defined in (11.13)-(11.18).

Theorem 11.8 *Let $\Lambda = \{f_s : s \in I \subseteq \mathbb{R}\}$ be a family of differentiable functions defined on $[a,b]$ such that the function $s \mapsto [z_0, z_1, z_2, z_3; f_s]$ is n-exponentially convex in the Jensen sense on I for every four mutually distinct points $z_0, z_1, z_2, z_3 \in [a,b]$. Let Ψ_i $(i = 1\text{-}6)$ be the linear functionals as defined in (11.13)-(11.18). Then the following statements hold:*

(i) *The function $s \mapsto \Psi_i(f_s)$ is n-exponentially convex in the Jensen sense on I and the matrix $\left[\Psi_i\left(f_{\frac{s_j+s_k}{2}}\right)\right]_{j,k=1}^{m}$ is positive semi-definite for all $m \in \mathbb{N}$, $m \leq n$ and $s_1, \ldots, s_m \in I$. Particularly,*

$$\det\left[\Psi_i\left(f_{\frac{s_j+s_k}{2}}\right)\right]_{j,k=1}^{m} \geq 0, \ \forall\, m \in \mathbb{N},\ m \leq n.$$

(ii) *If the function $s \mapsto \Psi_i(f_s)$ is continuous on I, then it is n-exponentially convex on I.*

Proof. The proof is analogous to the proof of Theorem 3.11 in [6]. □

The following corollary is an immediate consequence of Theorem 11.8.

Corollary 11.3 *Let $\Lambda = \{f_s : s \in I \subseteq \mathbb{R}\}$ be a family of differentiable functions defined on $[a,b]$ such that the function $s \mapsto [z_0, z_1, z_2, z_3; f_s]$ is exponentially convex in the Jensen sense on I for every four mutually distinct points $z_0, z_1, z_2, z_3 \in [a,b]$. Let Ψ_i $(i = 1\text{-}6)$ be the linear functionals as defined in* (11.13)-(11.18). *Then the following statements hold:*

(i) *The function $s \mapsto \Psi_i(f_s)$ is exponentially convex in the Jensen sense on I and the matrix $\left[\Psi_i\left(f_{\frac{s_j+s_k}{2}}\right)\right]_{j,k=1}^{m}$ is positive semi-definite for all $m \in \mathbb{N}$, $m \leq n$ and $s_1, \ldots, s_m \in I$. Particularly,*

$$\det\left[\Psi_i\left(f_{\frac{s_j+s_k}{2}}\right)\right]_{j,k=1}^{m} \geq 0, \ \forall\, m \in \mathbb{N},\ m \leq n.$$

(ii) *If the function $s \mapsto \Psi_i(f_s)$ is continuous on I, then it is exponentially convex on I.*

Corollary 11.4 *Let $\Lambda = \{f_s : s \in I \subseteq \mathbb{R}\}$ be a family of differentiable functions defined on $[a,b]$ such that the function $s \mapsto [z_0, z_1, z_2, z_3; f_s]$ is 2-exponentially convex in the Jensen sense on I for every four mutually distinct points $z_0, z_1, z_2, z_3 \in [a,b]$. Let Ψ_i $(i = 1\text{-}6)$ be the linear functionals as defined in* (11.13)-(11.18). *Further, assume that $\Psi_i(f_s)$ $(i = 1\text{-}6)$ is strictly positive for $f_s \in \Lambda$. Then the following statements hold:*

(i) *If the function $s \mapsto \Psi_i(f_s)$ is continuous on I, then it is 2-exponentially convex on I and so it is* log-*convex on I and for $\tilde{r}, s, \tilde{t} \in I$ such that $\tilde{r} < s < \tilde{t}$, we have*

$$[\Psi_i(f_s)]^{\tilde{t}-\tilde{r}} \leq [\Psi_i(f_{\tilde{r}})]^{\tilde{t}-s}\,[\Psi_i(f_{\tilde{t}})]^{s-\tilde{r}}, \qquad i \in \{1, \ldots, 6\}, \tag{11.20}$$

known as Lyapunov's inequality. If $\tilde{r} < \tilde{t} < s$ or $s < \tilde{r} < \tilde{t}$, then the reversed inequalities hold in (11.20).

(ii) *If the function $s \mapsto \Psi_i(f_s)$ is differentiable on I, then for every $s, q, u, v \in I$ such that $s \leq u$ and $q \leq v$, we have*

$$\mu_{s,q}(\Psi_i, \Lambda) \leq \mu_{u,v}(\Psi_i, \Lambda), \qquad i \in \{1, \ldots, 6\}, \tag{11.21}$$

where

$$\mu_{s,q}(\Psi_i, \Lambda) = \begin{cases} \left(\dfrac{\Psi_i(f_s)}{\Psi_i(f_q)}\right)^{\frac{1}{s-q}}, & s \neq q, \\ \exp\left(\dfrac{\frac{d}{ds}\Psi_i(f_s)}{\Psi_i(f_s)}\right), & s = q \end{cases} \tag{11.22}$$

for $f_s, f_q \in \Lambda$.

Proof. The proof is analogous to the proof of the Corollary 3.13 in [6]. □

Remark 11.3 Note that the results from Theorem 11.8, Corollary 11.3 and Corollary 11.4 still hold when two of the points $z_0, z_1, z_2, z_3 \in [a,b]$ coincide, say $z_1 = z_0$, for a family of differentiable functions f_s such that the function $s \mapsto [z_0, z_0, z_2, z_3; f_s]$ is n-exponentially convex in the Jensen sense (exponentially convex in the Jensen sense, log-convex in the Jensen sense on I), when three of the points $z_0, z_1, z_2, z_3 \in [a,b]$ coincide, say $z_2 = z_1 = z_0$, for a family of differentiable functions f_s such that the function $s \mapsto [z_0, z_0, z_0, z_3; f_s]$ is n-exponentially convex in the Jensen sense, when three of the points $z_0, z_1, z_2, z_3 \in [a,b]$ coincide again, say $z_2 = z_1 = z_0$, for a family of twice differentiable functions f_s such that the function $s \mapsto [z_0, z_0, z_0, z_3; f_s]$ is n-exponentially convex in the Jensen sense and furthermore, they still hold when all four points coincide for a family of thrice differentiable functions with the same property.

There are several families of functions which fulfil the conditions of Theorem 11.8, Corollaries 11.3 and 11.4, and Remark 11.3 and so the results of these theorem and corollaries can be applied for them. Here we present an example for such a family of functions and for more examples see [6] and [9].

Example 11.1 Consider the family of functions

$$\tilde{\Lambda} = \{f_s : (0,\infty) \to \mathbb{R} : s \in \mathbb{R}\}$$

defined by

$$f_s(x) = \begin{cases} \frac{x^s}{s(s-1)(s-2)}, & s \notin \{0,1,2\}, \\ \frac{1}{2}\log x, & s = 0, \\ -x\log x, & s = 1, \\ \frac{1}{2}x^2 \log x, & s = 2. \end{cases}$$

Here $\frac{d^3}{dx^3} f_s(x) = x^{s-3} = e^{(s-3)\ln x} > 0$, which shows that f_s is 3-convex for $x > 0$ and $s \mapsto \frac{d^3}{dx^3} f_s(x)$ is exponentially convex by definition.

In order to prove that $s \mapsto [z_0, z_1, z_2, z_3; f_s]$ is exponentially convex, it is enough to show that

$$\sum_{j,k=1}^n \varsigma_j \varsigma_k \left[z_0, z_1, z_2, z_3; f_{\frac{s_j+s_k}{2}}\right] = \left[z_0, z_1, z_2, z_3; \sum_{j,k=1}^n \varsigma_j \varsigma_k f_{\frac{s_j+s_k}{2}}\right] \geq 0, \tag{11.23}$$

for all $n \in \mathbb{N}$, $\varsigma_j, s_j \in \mathbb{R}$, $j \in \{1, \ldots, n\}$. By Definition 11.3, inequality (11.23) will hold if $\Gamma(x) := \sum_{j,k=1}^n \varsigma_j \varsigma_k f_{\frac{s_j+s_k}{2}}(x)$ is 3-convex. Since $s \mapsto \frac{d^3}{dx^3} f_s(x)$ is exponentially convex, that is

$$\sum_{j,k=1}^n \varsigma_j \varsigma_k f'''_{\frac{s_j+s_k}{2}} \geq 0, \qquad \forall\, n \in \mathbb{N},\ \varsigma_j, s_j \in \mathbb{R},\ j \in \{1, \ldots, n\},$$

which implies that Γ is 3-convex, inequality (11.23) is immediate. Now as $s \mapsto [z_0, z_1, z_2, z_3; f_s]$ is exponentially convex, $s \mapsto [z_0, z_1, z_2, z_3; f_s]$ is exponentially convex in the Jensen sense and by using Corollary 11.3, we have $s \mapsto \Psi_i(f_s)$ $(i = 1\text{-}6)$ is exponentially convex in the Jensen sense. Since these mappings are continuous, $s \mapsto \Psi_i(f_s)$

$(i = 1\text{-}6)$ is exponentially convex.

If $\tilde{r}, s, \tilde{t} \in \mathbb{R}$ are such that $\tilde{r} < s < \tilde{t}$, then from (11.20) we have

$$\Psi_i(f_s) \le [\Psi_i(f_{\tilde{r}})]^{\frac{\tilde{t}-s}{\tilde{t}-\tilde{r}}} [\Psi_i(f_{\tilde{t}})]^{\frac{s-\tilde{r}}{\tilde{t}-\tilde{r}}}, \qquad i \in \{1, \ldots, 6\}. \tag{11.24}$$

If $\tilde{r} < \tilde{t} < s$ or $s < \tilde{r} < \tilde{t}$, then the reversed inequality holds in (11.24).

Particularly, for $i = 1$ and $\tilde{r}, s, \tilde{t} \in \mathbb{R} \setminus \{0, 1, 2\}$ such that $\tilde{r} < s < \tilde{t}$, we have

$$\begin{aligned}
&\frac{1}{s(s-1)(s-2)} \sum_{k=2}^{n} \left(\frac{1}{H^s_{n,r,t}} \left(\left(\sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right)^s - \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right)^s \right) \right.\\
&\left. + a_k^s \left(C^s_{k-1,n,r,t} - C^s_{k,n,r,t} \right) + s a_k^{s-1} D_k \left(C^{s-1}_{k-1,n,r,t} - C^{s-1}_{k,n,r,t} \right) \right)\\
&\le \left[\frac{1}{\tilde{r}(\tilde{r}-1)(\tilde{r}-2)} \sum_{k=2}^{n} \left(\frac{1}{H^{\tilde{r}}_{n,r,t}} \left(\left(\sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right)^{\tilde{r}} - \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right)^{\tilde{r}} \right) \right. \right.\\
&\left. \left. + a_k^{\tilde{r}} \left(C^{\tilde{r}}_{k-1,n,r,t} - C^{\tilde{r}}_{k,n,r,t} \right) + \tilde{r} a_k^{\tilde{r}-1} D_k \left(C^{\tilde{r}-1}_{k-1,n,r,t} - C^{\tilde{r}-1}_{k,n,r,t} \right) \right) \right]^{\frac{\tilde{t}-s}{\tilde{t}-\tilde{r}}}\\
&\times \left[\frac{1}{\tilde{t}(\tilde{t}-1)(\tilde{t}-2)} \sum_{k=2}^{n} \left(\frac{1}{H^{\tilde{t}}_{n,r,t}} \left(\left(\sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right)^{\tilde{t}} - \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right)^{\tilde{t}} \right) \right. \right.\\
&\left. \left. + a_k^{\tilde{t}} \left(C^{\tilde{t}}_{k-1,n,r,t} - C^{\tilde{t}}_{k,n,r,t} \right) + \tilde{t} a_k^{\tilde{t}-1} D_k \left(C^{\tilde{t}-1}_{k-1,n,r,t} - C^{\tilde{t}-1}_{k,n,r,t} \right) \right) \right]^{\frac{s-\tilde{r}}{\tilde{t}-\tilde{r}}}, \ s \notin \{0, 1, 2\}.
\end{aligned}$$

In this case, $\mu_{s,q}(\Psi_i, \Lambda)$ $(i = 1\text{-}6)$ defined in (11.22) becomes

$$\mu_{s,q}(\Psi_i, \tilde{\Lambda}) = \begin{cases}
\left(\frac{\Psi_i(f_s)}{\Psi_i(f_q)} \right)^{\frac{1}{s-q}}, & s \ne q,\\
\exp\left(\frac{2\Psi_i(f_s f_0)}{\Psi_i(f_s)} - \frac{3s^2-6s+2}{s(s-1)(s-2)} \right), & s = q \notin \{0, 1, 2\},\\
\exp\left(\frac{\Psi_i(f_0^2)}{\Psi_i(f_0)} + \frac{3}{2} \right), & s = q = 0,\\
\exp\left(\frac{\Psi_i(f_0 f_1)}{\Psi_i(f_1)} \right), & s = q = 1,\\
\exp\left(\frac{\Psi_i(f_0 f_2)}{\Psi_i(f_2)} - \frac{3}{2} \right), & s = q = 2.
\end{cases}$$

In particular for $i = 1$, we have

$$\begin{aligned}
\Psi_1(f_s) &= \frac{1}{s(s-1)(s-2)} \sum_{k=2}^{n} \left(\frac{1}{H^s_{n,r,t}} \left(\left(\sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right)^s - \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right)^s \right) \right.\\
&\quad \left. + a_k^s \left(C^s_{k-1,n,r,t} - C^s_{k,n,r,t} + s a_k^{-1} D_k \left(C^{s-1}_{k-1,n,r,t} - C^{s-1}_{k,n,r,t} \right) \right) \right), s \notin \{0, 1, 2\},\\
\Psi_1(f_0) &= \frac{1}{2} \sum_{k=2}^{n} a_k^s \left(\log \left(\frac{C_{k-1,n,r,t} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t}}{C_{k,n,r,t} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t}} \right) + \frac{D_k}{a_k (k+r)^t C_{k-1,n,r,t} C_{k,n,r,t} H_{n,r,t}} \right),
\end{aligned}$$

$$
\begin{aligned}
\Psi_1(f_1) = & \frac{1}{H_{n,r,t}} \sum_{k=2}^{n} a_k^{s-1} \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \log \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right. \\
& \left. - \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \log \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) \right) \\
& + \sum_{k=2}^{n} a_k^s \left(C_{k,n,r,t} \log C_{k,n,r,t} - C_{k-1,n,r,t} \log C_{k-1,n,r,t} \right) \\
& + \sum_{k=2}^{n} a_k^{s-1} D_k \log \left(\frac{C_{k,n,r,t}}{C_{k-1,n,r,t}} \right), \\
2\Psi_1(f_2) = & \frac{1}{H_{n,r,t}^2} \sum_{k=2}^{n} a_k^{s-2} \left(\left(\sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right)^2 \log \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) \right. \\
& \left. - \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right)^2 \log \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right) \\
& + \sum_{k=2}^{n} a_k^s \left(C_{k-1,n,r,t}^2 \log \left(C_{k-1,n,r,t} \right) - C_{k,n,r,t}^2 \log \left(C_{k,n,r,t} \right) \right) \\
& - \sum_{k=2}^{n} a_k^{s-1} D_k \left(\frac{1}{(k+r)^t H_{n,r,t}} + 2C_{k,n,r,t} \log \left(C_{k,n,r,t} \right) \right. \\
& \left. - 2C_{k-1,n,r,t} \log \left(C_{k-1,n,r,t} \right) \right), \\
4\Psi_1\left(f_0{}^2\right) = & \sum_{k=2}^{n} a_k^s \left(\log^2 \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) - \log^2 \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right) \\
& + \sum_{k=2}^{n} a_k^s \left(\log^2 \left(C_{k-1,n,r,t} \right) - \log^2 \left(C_{k,n,r,t} \right) \right) \\
& - 2 \sum_{k=2}^{n} a_k^{s-1} D_k \left(\frac{\log \left(C_{k,n,r,t} \right)}{C_{k,n,r,t}} - \frac{\log \left(C_{k-1,n,r,t} \right)}{C_{k-1,n,r,t}} \right), \\
2\Psi_1(f_s f_0) = & \frac{1}{s(s-1)(s-2) H_{n,r,t}^s} \sum_{k=2}^{n} \left(\left(\sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right)^s \log \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) \right. \\
& \left. - \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right)^s \log \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right) \\
& + \frac{1}{s(s-1)(s-2)} \sum_{k=2}^{n} a_k^s \left(\left(C_{k-1,n,r,t}^s \log \left(C_{k-1,n,r,t} \right) - C_{k,n,r,t}^s \log \left(C_{k,n,r,t} \right) \right) \right. \\
& \left. + a_k^{-1} D_k \left(C_{k-1,n,r,t}^{s-1} \left(1 + s \log \left(C_{k-1,n,r,t} \right) \right) - C_{k,n,r,t}^{s-1} \left(1 + s \log \left(C_{k,n,r,t} \right) \right) \right) \right), \\
& s \notin \{0, 1, 2\},
\end{aligned}
$$

$$
\begin{aligned}
2\Psi_1(f_0 f_1) = {} & \frac{1}{H_{n,r,t}} \sum_{k=2}^{n} a_k^{s-1} \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \log^2 \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right. \\
& \left. - \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \log^2 \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) \right) \\
& + \sum_{k=2}^{n} a_k^s \left(C_{k,n,r,t} \log^2 \left(C_{k,n,r,t} \right) - C_{k-1,n,r,t} \log^2 \left(C_{k-1,n,r,t} \right) \right) \\
& + \sum_{k=2}^{n} a_k^{s-1} D_k \left(\left(2 + \log \left(C_{k,n,r,t} \right) \right) \log \left(C_{k,n,r,t} \right) \right. \\
& \left. - \left(2 + \log \left(C_{k-1,n,r,t} \right) \right) \log \left(C_{k-1,n,r,t} \right) \right)
\end{aligned}
$$

and

$$
\begin{aligned}
4\Psi_1(f_0 f_2) = {} & \frac{1}{H_{n,r,t}^2} \sum_{k=2}^{n} a_k^{s-2} \left(\left(\sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right)^2 \log^2 \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k} \frac{a_i}{(i+r)^t} \right) \right. \\
& \left. - \left(\sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right)^2 \log^2 \left(\frac{1}{a_k H_{n,r,t}} \sum_{i=1}^{k-1} \frac{a_i}{(i+r)^t} \right) \right) \\
& + \sum_{k=2}^{n} a_k^s \left(C_{k-1,n,r,t}^2 \log^2 \left(C_{k-1,n,r,t} \right) - C_{k,n,r,t}^2 \log^2 \left(C_{k,n,r,t} \right) \right) \\
& + 2 \sum_{k=2}^{n} a_k^{s-1} D_k \left(C_{k-1,n,r,t} \left(1 + \log \left(C_{k-1,n,r,t} \right) \right) \log \left(C_{k-1,n,r,t} \right) \right. \\
& \left. - C_{k,n,r,t} \left(1 + \log \left(C_{k,n,r,t} \right) \right) \log \left(C_{k,n,r,t} \right) \right).
\end{aligned}
$$

If Ψ_i $(i = 1\text{-}6)$ is positive, then Theorem 11.7 applied for $g = f_s \in \tilde{\Lambda}$ and $h = f_q \in \tilde{\Lambda}$ yields that there exists $\delta_i \in [a,b]$ such that

$$
\delta_i^{s-q} = \frac{\Psi_i(f_s)}{\Psi_i(f_q)}, \quad i \in \{1, \ldots, 6\}.
$$

Since the function $\delta_i \mapsto \delta_i^{s-q}$ $(i = 1\text{-}6)$ is invertible for $s \neq q$, we have

$$
\min\{a,b\} \leq \left(\frac{\Psi_i(f_s)}{\Psi_i(f_q)} \right)^{\frac{1}{s-q}} \leq \max\{a,b\}, \quad i \in \{1, \ldots, 6\},
$$

which together with the fact that $\mu_{s,q}(\Psi_i, \tilde{\Lambda})$ $(i = 1\text{-}6)$ is continuous, symmetric and monotonous (by (11.21)) shows that $\mu_{s,q}\left(\Psi_i, \tilde{\Lambda}\right)$ $(i = 1\text{-}6)$ is a mean.

11.3 Shannon and Zipf-Mandelbrot entropies and related results

The results presented in this section are given in [8]. This section is organized as follows: in Section 11.3.1 and 11.3.2, we present some interesting results related to Shannon and Zipf-Mandelbrot entropies respectively. In Section 11.3.3, we define linear functionals as the non-negative differences of the obtained inequalities and present mean value theorems for the linear functionals. In Section 11.3.4, we present the properties of the functionals, such as n-exponential and logarithmic convexity. Finally, we give an example of the family of functions for which the results can be applied.

Remark 11.4 log denotes the logarithmic function and throughout this section we consider the base b of logarithm is greater than 1.

11.3.1 Inequalities related to Shannon entropy

In our first main result, we will use the following lemma:

Lemma 11.1 *(i) If $p_i \in \mathbb{R}$ such that $p_i > 0$ $(1 \leq i \leq n)$ and if the sequence $\{a_i\}_{i=1}^n \subset \mathbb{R}$ is non-increasing, then it is non-increasing in weighted mean.*

(ii) If $p_i \in \mathbb{R}$ such that $p_i > 0$ $(1 \leq i \leq n)$ and if the sequence $\{a_i\}_{i=1}^n \subset \mathbb{R}$ is non-decreasing, then it is non-decreasing in weighted mean.

Proof.

(i) Simple calculations reveal that

$$\frac{1}{P_k}\sum_{i=1}^{k} p_i a_i - \frac{1}{P_{k+1}}\sum_{i=1}^{k+1} p_i a_i = \frac{p_{k+1}}{P_k P_{k+1}}\left(\sum_{i=1}^{k} p_i a_i - P_k a_{k+1}\right).$$

As $a_1 \geq \ldots \geq a_n$ and $p_i > 0$, $i \in \{1,\ldots,n\}$, we have

$$\begin{aligned} p_1 a_1 &\geq p_1 a_{k+1}, \\ &\vdots \\ p_k a_k &\geq p_k a_{k+1}. \end{aligned}$$

On adding the above inequalities we have $\sum_{i=1}^{k} p_i a_i - P_k a_{k+1} \geq 0$, which combined together with $\frac{p_{k+1}}{P_k P_{k+1}} > 0$ yields that $\frac{1}{P_k}\sum_{i=1}^{k} p_i a_i \geq \frac{1}{P_{k+1}}\sum_{i=1}^{k+1} p_i a_i$.

(ii) The proof is analogous to the proof of (i).

□

Our first main result states that:

Theorem 11.9 *Let $p_k \in \mathbb{R}$ such that $p_k > 0$ $(1 \le k \le n)$ and let $f : [a,b] \to \mathbb{R}$ be a Wright-convex function.*

(a) Let $0 < p_k < 1$ $(1 \le k \le n)$ and let $S(\mathbf{p})$, $p_1 \log\left(\frac{1}{p_1}\right)$, $P_k \log\left(\frac{1}{p_k}\right)$, $P_{k-1} \log\left(\frac{1}{p_k}\right)$ $\in [a,b]$ for all $k \in \{2,\dots,n\}$.

(i) If the sequence $\{p_k\}_{k=1}^n$ is non-increasing, then

$$f(S(\mathbf{p})) \le f\left(p_1 \log\left(\frac{1}{p_1}\right)\right) + \sum_{k=2}^{n}\left[f\left(P_k \log\left(\frac{1}{p_k}\right)\right) - f\left(P_{k-1} \log\left(\frac{1}{p_k}\right)\right)\right]. \tag{11.25}$$

(ii) If the sequence $\{p_k\}_{k=1}^n$ is non-decreasing, then

$$f(S(\mathbf{p})) \ge f\left(p_1 \log\left(\frac{1}{p_1}\right)\right) + \sum_{k=2}^{n}\left[f\left(P_k \log\left(\frac{1}{p_k}\right)\right) - f\left(P_{k-1} \log\left(\frac{1}{p_k}\right)\right)\right]. \tag{11.26}$$

(b) Let $p_k \ge 1$ $(1 \le k \le n)$ and let $-S(\mathbf{p})$, $p_1 \log p_1$, $P_k \log p_k$, $P_{k-1} \log p_k \in [a,b]$ for all $k \in \{2,\dots,n\}$.

(i) If the sequence $\{p_k\}_{k=1}^n$ is non-increasing, then

$$f(-S(\mathbf{p})) \ge f(p_1 \log p_1) + \sum_{k=2}^{n}\left[f(P_k \log p_k) - f(P_{k-1} \log p_k)\right]. \tag{11.27}$$

(ii) If the sequence $\{p_k\}_{k=1}^n$ is non-decreasing, then

$$f(-S(\mathbf{p})) \le f(p_1 \log p_1) + \sum_{k=2}^{n}\left[f(P_k \log p_k) - f(P_{k-1} \log p_k)\right]. \tag{11.28}$$

If f is Wright-concave, then the reversed inequalities hold in (11.25) - (11.28).

Proof.

(a) (i) As $p_1 \ge p_2 \ge \ldots \ge p_n$ and $b > 1$, the sequence $\left\{\log\left(\frac{1}{p_k}\right)\right\}_{k=1}^n$ is non-decreasing. By Lemma 11.1 (ii), the sequence $\left\{\log\left(\frac{1}{p_k}\right)\right\}_{k=1}^n$ is non-decreasing in weighted mean and hence by using Theorem 11.3(ii) for $a_k = \log\left(\frac{1}{p_k}\right)$ such that $0 < p_k < 1$ $(1 \le k \le n)$, the result is immediate.

(ii) The idea of the proof is the same as discussed in (i).

(b) (i) As $p_1 \geq p_2 \geq \ldots \geq p_n$ and $b > 1$, the sequence $\{\log p_k\}_{k=1}^n$ is non-increasing. By Lemma 11.1 (i), the sequence $\{\log p_k\}_{k=1}^n$ is non-increasing in weighted mean and hence by taking $a_k = \log p_k$ with $p_k \geq 1$ $(1 \leq k \leq n)$ in Theorem 11.3(i), the result is immediate.

(ii) The idea of the proof is the same as discussed in (i).

□

Since the class of convex (concave) functions is properly contained in the class of Wright-convex (Wright-concave) functions, the following corollary is immediate:

Corollary 11.5 *Let $p_k \in \mathbb{R}$ such that $p_k > 0$ $(1 \leq k \leq n)$ and let $f : [a,b] \to \mathbb{R}$ be a convex function.*

(a) Let all the conditions of Theorem 11.9(a) hold.

(i) If the sequence $\{p_k\}_{k=1}^n$ is non-increasing, then (11.25) holds.

(ii) If the sequence $\{p_k\}_{k=1}^n$ is non-decreasing, then (11.26) holds.

(b) Let all the conditions of Theorem 11.9(b) hold.

(i) If the sequence $\{p_k\}_{k=1}^n$ is non-increasing, then (11.27) holds.

(ii) If the sequence $\{p_k\}_{k=1}^n$ is non-decreasing, then (11.28) holds.

If f is concave, then the reversed inequalities hold in (11.25) - (11.28).

An application of Corollary 11.5 is given as follows:

Corollary 11.6 *Let $f(x) = x^s$, where $x \in (0,\infty)$ and $s \in \mathbb{R}$. Let $p_k \in \mathbb{R}$ such that $p_k > 0$ $(1 \leq k \leq n)$.*

(a) Let $0 < p_k < 1$ $(1 \leq k \leq n)$ and let $s < 0$ or $s > 1$.

(i) If the sequence $\{p_k\}_{k=1}^n$ is non-increasing, then

$$(S(\mathbf{p}))^s \leq \left(p_1 \log\left(\frac{1}{p_1}\right)\right)^s + \sum_{k=2}^n \left(\log\left(\frac{1}{p_k}\right)\right)^s \left(P_k^s - P_{k-1}^s\right). \tag{11.29}$$

(ii) If the sequence $\{p_k\}_{k=1}^n$ is non-decreasing, then

$$(S(\mathbf{p}))^s \geq \left(p_1 \log\left(\frac{1}{p_1}\right)\right)^s + \sum_{k=2}^n \left(\log\left(\frac{1}{p_k}\right)\right)^s \left(P_k^s - P_{k-1}^s\right). \tag{11.30}$$

(b) Let $p_k \geq 1$ $(1 \leq k \leq n)$ and let $s < 0$ or $s > 1$.

(i) If the sequence $\{p_k\}_{k=1}^n$ is non-increasing, then

$$(-S(\mathbf{p}))^s \geq (p_1 \log p_1)^s + \sum_{k=2}^n (\log p_k)^s \left(P_k^s - P_{k-1}^s\right). \tag{11.31}$$

(ii) If the sequence $\{p_k\}_{k=1}^n$ *is non-decreasing, then*

$$(-S(\mathbf{p}))^s \leq (p_1 \log p_1)^s + \sum_{k=2}^{n} (\log p_k)^s \left(P_k^s - P_{k-1}^s\right). \tag{11.32}$$

If $0 < s < 1$*, then the reversed inequalities hold in* (11.29) - (11.32).

11.3.2 Inequalities related to Zipf-Mandelbrot entropy

The aim of this section is to present some interesting results by using Zipf-Mandelbrot entropy.

The first main result of this section states that:

Theorem 11.10 *Let* $Z(r,t,H_{n,r,t})$ *be the Zipf-Mandelbrot entropy,* $C_{k,n,r,t}$ *be the cumulative distribution function and* $f:[a,b]\to\mathbb{R}$ *be a Wright-convex function.*

(i) Let $0 < \frac{1}{(k+r)^t H_{n,r,t}} < 1$.

If $Z(r,t,H_{n,r,t})$, $\log\left((1+r)^t H_{n,r,t}\right)^{\frac{1}{(1+r)^t H_{n,r,t}}}$, $\log\left((k+r)^t H_{n,r,t}\right)^{C_{k,n,r,t}}$, $\log\left((k+r)^t H_{n,r,t}\right)^{C_{k-1,n,r,t}} \in [a,b]$ *for all* $k\in\{2,\ldots,n\}$*, then*

$$\begin{aligned} f\left(Z(r,t,H_{n,r,t})\right) \leq\ & f\left(\log\left((1+r)^t H_{n,r,t}\right)^{\frac{1}{(1+r)^t H_{n,r,t}}}\right) \\ & + \sum_{k=2}^{n} f\left(\log\left((k+r)^t H_{n,r,t}\right)^{C_{k,n,r,t}}\right) \\ & - \sum_{k=2}^{n} f\left(\log\left((k+r)^t H_{n,r,t}\right)^{C_{k-1,n,r,t}}\right). \end{aligned} \tag{11.33}$$

(ii) Let $(k+r)^t H_{n,r,t} \leq 1$.

If $-Z(r,t,H_{n,r,t})$, $\log\left((1+r)^t H_{n,r,t}\right)^{\frac{-1}{(1+r)^t H_{n,r,t}}}$, $\log\left((k+r)^t H_{n,r,t}\right)^{-C_{k,n,r,t}}$, $\log\left((k+r)^t H_{n,r,t}\right)^{-C_{k-1,n,r,t}} \in [a,b]$ *for all* $k\in\{2,\ldots,n\}$*, then*

$$\begin{aligned} f\left(-Z(r,t,H_{n,r,t})\right) \geq\ & f\left(\log\left((1+r)^t H_{n,r,t}\right)^{\frac{-1}{(1+r)^t H_{n,r,t}}}\right) \\ & + \sum_{k=2}^{n} f\left(\log\left((k+r)^t H_{n,r,t}\right)^{-C_{k,n,r,t}}\right) \\ & - \sum_{k=2}^{n} f\left(\log\left((k+r)^t H_{n,r,t}\right)^{-C_{k-1,n,r,t}}\right). \end{aligned} \tag{11.34}$$

If f *is Wright-concave, then the reversed inequalities hold in* (11.33) *and* (11.34).

Proof. It is easy to see that the sequence $\left\{p_k = \frac{1}{(k+r)^t H_{n,r,t}}\right\}_{k=1}^{n}$ is non-increasing over $k\in\{1,\ldots,n\}$.

(i) Take $p_k = \frac{1}{(k+r)^t H_{n,r,t}}$ in Theorem 11.9(a) (i), the result is immediate.

(ii) The idea of the proof is the same as discussed in (i) but here we apply Theorem 11.9(b) (i) instead of Theorem 11.9(a) (i).

□

Corollary 11.7 *Let $Z(r,t,H_{n,r,t})$ be the Zipf-Mandelbrot entropy, $C_{k,n,r,t}$ be the cumulative distribution function and $f:[a,b]\to\mathbb{R}$ be a convex function.*

(i) If all the conditions of Theorem 11.10(i) *hold, then we have* (11.33).

(ii) If all the conditions of Theorem 11.10(ii) *hold, then we have* (11.34).

If f is concave, then the reversed inequalities hold in (11.33) *and* (11.34).

11.3.3 Linear functionals and mean value theorems

Consider the inequalities (11.25), (11.27) and (11.33) and define linear functionals as follows:

$$\begin{aligned}\Phi_1(f) &= -f(S(\mathbf{p}))+f\left(p_1\log\left(\frac{1}{p_1}\right)\right)\\ &\quad+\sum_{k=2}^{n}\left[f\left(P_k\log\left(\frac{1}{p_k}\right)\right)-f\left(P_{k-1}\log\left(\frac{1}{p_k}\right)\right)\right], \qquad (11.35)\\ \Phi_2(f) &= f(-S(\mathbf{p}))-f(p_1\log p_1)-\sum_{k=2}^{n}\left[f(P_k\log p_k)-f(P_{k-1}\log p_k)\right] \qquad (11.36)\end{aligned}$$

and

$$\begin{aligned}\Phi_3(f) &= -f(Z(r,t,H_{n,r,t}))+f\left(\log\left((1+r)^t H_{n,r,t}\right)^{\frac{1}{(1+r)^t H_{n,r,t}}}\right)\\ &\quad+\sum_{k=2}^{n}\left[f\left(\log\left((k+r)^t H_{n,r,t}\right)^{C_{k,n,r,t}}\right)-f\left(\log\left((k+r)^t H_{n,r,t}\right)^{C_{k-1,n,r,t}}\right)\right]. \qquad (11.37)\end{aligned}$$

If f is a convex function defined on $[a,b]$ and if the sequence $\{p_k\}_{k=1}^{n}\subset\mathbb{R}$ is non-increasing, then Corollary 11.5$(a)(i)$ and Corollary 11.5$(b)(i)$ imply that $\Phi_1(f)\geq 0$ and $\Phi_2(f)\geq 0$ respectively. Moreover, if $Z(r,t,H_{n,r,t})$ is the Zipf-Mandelbrot entropy, $C_{k,n,r,t}$ is the cumulative distribution function and if $f:[a,b]\to\mathbb{R}$ is a convex function, then Corollary 11.7(i) implies that $\Phi_3(f)\geq 0$.

Now we present mean value theorems for the functional Φ_i $(i=1\text{-}3)$. Lagrange-type mean value theorem related to Φ_i $(i=1\text{-}3)$ states that :

Theorem 11.11 *Let $f : [a,b] \to \mathbb{R}$ be such that $f \in C^2([a,b])$ and let Φ_1, Φ_2 and Φ_3 be the linear functionals as defined in* (11.35), (11.36) *and* (11.37) *respectively. Then there exists $\xi_1, \xi_2, \xi_3 \in [a,b]$ such that*

$$\Phi_i(f) = \frac{f''(\xi_i)}{2}\Phi_i(f_0), \qquad i \in \{1,2,3\},$$

where $f_0(x) = x^2$.

Proof. The proof is analogous to the proof of Theorem 2.7 given in [5] (see also Theorem 2.2 in [14]). □

The following theorem is a new analogue of the classical Cauchy mean value theorem, related to Φ_i $(i = 1\text{-}3)$.

Theorem 11.12 *Let $f, g : [a,b] \to \mathbb{R}$ be such that $f, g \in C^2([a,b])$ and let Φ_1, Φ_2 and Φ_3 be the linear functionals as defined in* (11.35), (11.36) *and* (11.37) *respectively. Then there exist $\xi_1, \xi_2, \xi_3 \in [a,b]$ such that*

$$\frac{\Phi_i(f)}{\Phi_i(g)} = \frac{f''(\xi_i)}{g''(\xi_i)}, \qquad i \in \{1,2,3\}, \tag{11.38}$$

provided that the denominators are non-zero.

Proof. The proof is analogous to the proof of Theorem 2.8 given in [5] (see also Theorem 2.4 in [14]). □

Remark 11.5 (i) By taking $f(x) = x^s$ and $g(x) = x^q$ in (11.38), where $s, q \in \mathbb{R} \setminus \{0,1\}$ are such that $s \neq q$, we have

$$\xi_i^{s-q} = \frac{q(q-1)\Phi_i(x^s)}{s(s-1)\Phi_i(x^q)}, \qquad i \in \{1,2,3\}.$$

(ii) If the inverse of the function f''/g'' exists, then (11.38) implies that

$$\xi_i = \left(\frac{f''}{g''}\right)^{-1}\left(\frac{\Phi_i(f)}{\Phi_i(g)}\right), \qquad i \in \{1,2,3\}.$$

11.3.4 *n*-exponential convexity and log-convexity

Let I be an open interval in $\mathbb{R}$.

Now we study the n-exponential convexity and log-convexity of the functions associated with the linear functionals Φ_i $(i = 1\text{-}3)$ as defined in (11.35)-(11.37).

Theorem 11.13 *Let $\Omega = \{f_s : s \in I \subseteq \mathbb{R}\}$ be a family of functions defined on $[a,b]$ such that the function $s \mapsto [z_0, z_1, z_2; f_s]$ is n-exponentially convex in the Jensen sense on I for every three mutually distinct points $z_0, z_1, z_2 \in [a,b]$. Let Φ_i $(i = 1\text{-}3)$ be the linear functionals as defined in* (11.35)-(11.37). *Then the following statements hold:*

(i) *The function* $s \mapsto \Phi_i(f_s)$ *is* n*-exponentially convex in the Jensen sense on* I *and the matrix* $\left[\Phi_i\left(f_{\frac{s_j+s_k}{2}}\right)\right]_{j,k=1}^{m}$ *is positive semi-definite for all* $m \in \mathbb{N}$, $m \leq n$ *and* $s_1, \ldots, s_m \in I$. *Particularly,*

$$\det\left[\Phi_i\left(f_{\frac{s_j+s_k}{2}}\right)\right]_{j,k=1}^{m} \geq 0, \ \forall\, m \in \mathbb{N},\, m \leq n.$$

(ii) *If the function* $s \mapsto \Phi_i(f_s)$ *is continuous on* I, *then it is* n*-exponentially convex on* I.

Proof. The idea of the proof is the same as that of the proof of Theorem 9 in [6]. □

The following corollary is an immediate consequence of Theorem 11.13.

Corollary 11.8 *Let* $\Omega = \{f_s : s \in I \subseteq \mathbb{R}\}$ *be a family of functions defined on* $[a,b]$ *such that the function* $s \mapsto [z_0, z_1, z_2; f_s]$ *is exponentially convex in the Jensen sense on* I *for every three mutually distinct points* $z_0, z_1, z_2 \in [a,b]$. *Let* Φ_i $(i = 1\text{-}3)$ *be the linear functionals as defined in* (11.35)-(11.37). *Then the following statements hold:*

(i) *The function* $s \mapsto \Phi_i(f_s)$ *is exponentially convex in the Jensen sense on* I *and the matrix* $\left[\Phi_i\left(f_{\frac{s_j+s_k}{2}}\right)\right]_{j,k=1}^{m}$ *is positive semi-definite for all* $m \in \mathbb{N}$, $m \leq n$ *and* $s_1, \ldots, s_m \in I$. *Particularly,*

$$\det\left[\Phi_i\left(f_{\frac{s_j+s_k}{2}}\right)\right]_{j,k=1}^{m} \geq 0, \ \forall\, m \in \mathbb{N},\, m \leq n.$$

(ii) *If the function* $s \mapsto \Phi_i(f_s)$ *is continuous on* I, *then it is exponentially convex on* I.

Corollary 11.9 *Let* $\Omega = \{f_s : s \in I \subseteq \mathbb{R}\}$ *be a family of functions defined on* $[a,b]$ *such that the function* $s \mapsto [z_0, z_1, z_2; f_s]$ *is 2-exponentially convex in the Jensen sense on* I *for every three mutually distinct points* $z_0, z_1, z_2 \in [a,b]$. *Let* Φ_i $(i = 1\text{-}3)$ *be the linear functionals as defined in* (11.35)-(11.37). *Further, assume that* $\Phi_i(f_s)$ $(i = 1\text{-}3)$ *is strictly positive for* $f_s \in \Omega$. *Then the following statements hold:*

(i) *If the function* $s \mapsto \Phi_i(f_s)$ *is continuous on* I, *then it is 2-exponentially convex on* I *and so it is* log-convex *on* I *and for* $\tilde{r}, s, \tilde{t} \in I$ *such that* $\tilde{r} < s < \tilde{t}$, *we have*

$$[\Phi_i(f_s)]^{\tilde{t}-\tilde{r}} \leq [\Phi_i(f_{\tilde{r}})]^{\tilde{t}-s} [\Phi_i(f_{\tilde{t}})]^{s-\tilde{r}}, \qquad i \in \{1,2,3\}, \tag{11.39}$$

known as Lyapunov's inequality. If $\tilde{r} < \tilde{t} < s$ *or* $s < \tilde{r} < \tilde{t}$, *then the reversed inequalities hold in* (11.39).

(ii) *If the function* $s \mapsto \Phi_i(f_s)$ *is differentiable on* I, *then for every* $s, q, u, v \in I$ *such that* $s \leq u$ *and* $q \leq v$, *we have*

$$\mu_{s,q}(\Phi_i, \Omega) \leq \mu_{u,v}(\Phi_i, \Omega), \qquad i \in \{1,2,3\}, \tag{11.40}$$

where

$$\mu_{s,q}(\Phi_i,\Omega)=\begin{cases}\left(\dfrac{\Phi_i(f_s)}{\Phi_i(f_q)}\right)^{\frac{1}{s-q}}, & s\neq q,\\ \exp\left(\dfrac{\frac{d}{ds}\Phi_i(f_s)}{\Phi_i(f_s)}\right), & s=q\end{cases} \tag{11.41}$$

for $f_s,f_q\in\Omega$.

Proof. The idea of the proof is the same as that of the proof of Corollary 5 given in [6]. □

Remark 11.6 Note that the results from Theorem 11.13, Corollary 11.8 and Corollary 11.9 still hold when two of the points $z_0,z_1,z_2\in[a,b]$ coincide, say $z_1=z_0$, for a family of differentiable functions f_s such that the function $s\mapsto[z_0,z_1,z_2;f_s]$ is n-exponentially convex in the Jensen sense (exponentially convex in the Jensen sense, log-convex in the Jensen sense on I); and furthermore, they still hold when all three points coincide for a family of twice differentiable functions with the same property.

There are several families of functions which fulfil the conditions of Theorem 11.13, Corollaries 11.8 and 11.9, and Remark 11.6 and so the results of these theorem and corollaries can be applied for them. Here we present an example for such a family of functions and for more examples see [9].

Example 11.2 Consider the family of functions

$$\tilde{\Omega}=\{f_s:(0,\infty)\to\mathbb{R}:s\in\mathbb{R}\}$$

defined by

$$f_s(x)=\begin{cases}\frac{x^s}{s(s-1)}, & s\notin\{0,1\},\\ -\log x, & s=0,\\ x\log x, & s=1.\end{cases}$$

Here $\frac{d^2}{dx^2}f_s(x)=x^{s-2}=e^{(s-2)\log x}>0$, which shows that f_s is convex for $x>0$ and $s\mapsto\frac{d^2}{dx^2}f_s(x)$ is exponentially convex by definition.

In order to prove that the function $s\mapsto[z_0,z_1,z_2;f_s]$ is exponentially convex, it is enough to show that

$$\textstyle\sum_{j,k=1}^{n}\varsigma_j\varsigma_k\left[z_0,z_1,z_2;f_{\frac{s_j+s_k}{2}}\right]=\left[z_0,z_1,z_2;\sum_{j,k=1}^{n}\varsigma_j\varsigma_k f_{\frac{s_j+s_k}{2}}\right]\geq 0, \tag{11.42}$$

for all $n\in\mathbb{N}$, $\varsigma_j,s_j\in\mathbb{R}$, $j\in\{1,\ldots,n\}$. By Definition 11.3, inequality (11.42) will hold if $\Xi(x):=\sum_{j,k=1}^{n}\varsigma_j\varsigma_k f_{\frac{s_j+s_k}{2}}(x)$ is convex. Since $s\mapsto\frac{d^2}{dx^2}f_s(x)$ is exponentially convex, that is

$$\sum_{j,k=1}^{n}\varsigma_j\varsigma_k f''_{\frac{s_j+s_k}{2}}\geq 0,\qquad \forall\, n\in\mathbb{N},\ \varsigma_j,s_j\in\mathbb{R},\ j\in\{1,\ldots,n\},$$

which shows the convexity of Ξ, inequality (11.42) is immediate. Now as the function $s\mapsto[z_0,z_1,z_2;f_s]$ is exponentially convex, $s\mapsto[z_0,z_1,z_2;f_s]$ is exponentially convex in the

Jensen sense and by using Corollary 11.8, we have $s \mapsto \Phi_i(f_s)$ $(i = 1\text{-}3)$ is exponentially convex in the Jensen sense. Since these mappings are continuous, $s \mapsto \Phi_i(f_s)$ $(i = 1\text{-}3)$ is exponentially convex.

If $\tilde{r}, s, \tilde{t} \in \mathbb{R}$ are such that $\tilde{r} < s < \tilde{t}$, then from (11.39) we have

$$\Phi_i(f_s) \le [\Phi_i(f_{\tilde{r}})]^{\frac{\tilde{t}-s}{\tilde{t}-\tilde{r}}} [\Phi_i(f_{\tilde{t}})]^{\frac{s-\tilde{r}}{\tilde{t}-\tilde{r}}}, \qquad i \in \{1,2,3\}. \tag{11.43}$$

If $\tilde{r} < \tilde{t} < s$ or $s < \tilde{r} < \tilde{t}$, then the reversed inequality holds in (11.43).

Particularly, for $i \in \{1,2,3\}$ and $\tilde{r}, s, \tilde{t} \in \mathbb{R} \setminus \{0,1\}$ such that $\tilde{r} < s < \tilde{t}$, we have

$$\begin{aligned}
&\frac{-S^s(\mathbf{p}) + \left(p_1 \log\left(\frac{1}{p_1}\right)\right)^s + \sum_{k=2}^n \left(\log\left(\frac{1}{p_k}\right)\right)^s \left(P_k^s - P_{k-1}^s\right)}{s(s-1)} \\
&\le \left[\frac{-S^{\tilde{r}}(\mathbf{p}) + \left(p_1 \log\left(\frac{1}{p_1}\right)\right)^{\tilde{r}} + \sum_{k=2}^n \left(\log\left(\frac{1}{p_k}\right)\right)^{\tilde{r}} \left(P_k^{\tilde{r}} - P_{k-1}^{\tilde{r}}\right)}{\tilde{r}(\tilde{r}-1)}\right]^{\frac{\tilde{t}-s}{\tilde{t}-\tilde{r}}} \\
&\times \left[\frac{-S^{\tilde{t}}(\mathbf{p}) + \left(p_1 \log\left(\frac{1}{p_1}\right)\right)^{\tilde{t}} + \sum_{k=2}^n \left(\log\left(\frac{1}{p_k}\right)\right)^{\tilde{t}} \left(P_k^{\tilde{t}} - P_{k-1}^{\tilde{t}}\right)}{\tilde{t}(\tilde{t}-1)}\right]^{\frac{s-\tilde{r}}{\tilde{t}-\tilde{r}}},
\end{aligned}$$

$$\begin{aligned}
&\frac{(-S(\mathbf{p}))^s - (p_1 \log p_1)^s - \sum_{k=2}^n (\log p_k)^s \left(P_k^s - P_{k-1}^s\right)}{s(s-1)} \\
&\le \left[\frac{(-S(\mathbf{p}))^{\tilde{r}} - (p_1 \log p_1)^{\tilde{r}} - \sum_{k=2}^n (\log p_k)^{\tilde{r}} \left(P_k^{\tilde{r}} - P_{k-1}^{\tilde{r}}\right)}{\tilde{r}(\tilde{r}-1)}\right]^{\frac{\tilde{t}-s}{\tilde{t}-\tilde{r}}} \\
&\times \left[\frac{(-S(\mathbf{p}))^{\tilde{t}} - (p_1 \log p_1)^{\tilde{t}} - \sum_{k=2}^n (\log p_k)^{\tilde{t}} \left(P_k^{\tilde{t}} - P_{k-1}^{\tilde{t}}\right)}{\tilde{t}(\tilde{t}-1)}\right]^{\frac{s-\tilde{r}}{\tilde{t}-\tilde{r}}}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{s(s-1)} \Bigg[-Z^s(r,t,H_{n,r,t}) + \left(\log\left((1+r)^t H_{n,r,t}\right)^{\frac{1}{(1+r)^t H_{n,r,t}}}\right)^s \\
&+ \sum_{k=2}^n \left[\left(\log\left((k+r)^t H_{n,r,t}\right)^{C_{k,n,r,t}}\right)^s - \left(\log\left((k+r)^t H_{n,r,t}\right)^{C_{k-1,n,r,t}}\right)^s\right]\Bigg] \\
&\le \left(\frac{1}{\tilde{r}(\tilde{r}-1)}\right)^{\frac{\tilde{t}-s}{\tilde{t}-\tilde{r}}} \Bigg[-Z^{\tilde{r}}(r,t,H_{n,r,t}) + \left(\log\left((1+r)^t H_{n,r,t}\right)^{\frac{1}{(1+r)^t H_{n,r,t}}}\right)^{\tilde{r}} \\
&+ \sum_{k=2}^n \left[\left(\log\left((k+r)^t H_{n,r,t}\right)^{C_{k,n,r,t}}\right)^{\tilde{r}} - \left(\log\left((k+r)^t H_{n,r,t}\right)^{C_{k-1,n,r,t}}\right)^{\tilde{r}}\right]\Bigg]^{\frac{\tilde{t}-s}{\tilde{t}-\tilde{r}}}
\end{aligned}$$

$$\times\left(\frac{1}{\tilde{t}\,(\tilde{t}-1)}\right)^{\frac{s-\tilde{r}}{\tilde{t}-\tilde{r}}}\left[-Z^{\tilde{t}}\,(r,t,H_{n,r,t})+\left(\log\left((1+r)^{t}\,H_{n,r,t}\right)^{\frac{1}{(1+r)^{t}H_{n,r,t}}}\right)^{\tilde{t}}\right.$$
$$\left.+\sum_{k=2}^{n}\left[\left(\log\left((k+r)^{t}\,H_{n,r,t}\right)^{C_{k,n,r,t}}\right)^{\tilde{t}}-\left(\log\left((k+r)^{t}\,H_{n,r,t}\right)^{C_{k-1,n,r,t}}\right)^{\tilde{t}}\right]\right]^{\frac{s-\tilde{r}}{\tilde{t}-\tilde{r}}}.$$

In this case, $\mu_{s,q}\,(\Phi_i,\Omega)$ $(i=1\text{-}3)$ defined in (11.41) becomes

$$\mu_{s,q}\left(\Phi_i,\tilde{\Omega}\right)=\begin{cases}\left(\frac{\Phi_i(f_s)}{\Phi_i(f_q)}\right)^{\frac{1}{s-q}}, & s\neq q,\\ \exp\left(\frac{1-2s}{s(s-1)}-\frac{\Phi_i(f_0f_s)}{\Phi_i(f_s)}\right), & s=q\notin\{0,1\},\\ \exp\left(1-\frac{\Phi_i(f_0^2)}{2\Phi_i(f_0)}\right), & s=q=0,\\ \exp\left(-1-\frac{\Phi_i(f_0f_1)}{2\Phi_i(f_1)}\right), & s=q=1.\end{cases}$$

In particular for $i=1$, we have

$$\Phi_1\,(f_s)=\frac{1}{s\,(s-1)}\left(-S^{s}\,(\mathbf{p})+p_1^{s}\log^{s}\left(\frac{1}{p_1}\right)+\sum_{k=2}^{n}\log^{s}\left(\frac{1}{p_k}\right)\left(P_k^{s}-P_{k-1}^{s}\right)\right),\ s\notin\{0,1\},$$

$$\Phi_1\,(f_0)=\log\left(\frac{S\,(\mathbf{p})}{p_1\log\left(\frac{1}{p_1}\right)}\right)+\sum_{k=2}^{n}\log\left(\frac{P_{k-1}}{P_k}\right),$$

$$\Phi_1\,(f_1)=\log\left(\frac{\left(p_1\log\left(\frac{1}{p_1}\right)\right)^{p_1\log\left(\frac{1}{p_1}\right)}}{(S\,(\mathbf{p}))^{S(\mathbf{p})}}\right)+\sum_{k=2}^{n}\log\left(\frac{\left(P_k\log\left(\frac{1}{p_k}\right)\right)^{P_k\log\left(\frac{1}{p_k}\right)}}{\left(P_{k-1}\log\left(\frac{1}{p_k}\right)\right)^{P_{k-1}\log\left(\frac{1}{p_k}\right)}}\right),$$

$$\Phi_1\left(f_0^2\right)=\sum_{k=2}^{n}\left[\log^2\left(P_k\log\left(\frac{1}{p_k}\right)\right)-\log^2\left(P_{k-1}\log\left(\frac{1}{p_k}\right)\right)\right]+\log^2\left(p_1\log\left(\frac{1}{p_1}\right)\right)$$
$$-\log^2\left(S\,(\mathbf{p})\right),$$

$$\Phi_1\,(f_0f_1)=S\,(\mathbf{p})\log^2\left(S\,(\mathbf{p})\right)-p_1\log\left(\frac{1}{p_1}\right)\log^2\left(p_1\log\left(\frac{1}{p_1}\right)\right)$$
$$-\sum_{k=2}^{n}\log\left(\frac{1}{p_k}\right)\left[P_k\log^2\left(P_k\log\left(\frac{1}{p_k}\right)\right)-P_{k-1}\log^2\left(P_{k-1}\log\left(\frac{1}{p_k}\right)\right)\right]$$

and

$$\Phi_1\,(f_0f_s)=\frac{1}{s\,(s-1)}\left(\log\left(\frac{(S\,(\mathbf{p}))^{(S(\mathbf{p}))^{s}}}{\left(p_1\log\left(\frac{1}{p_1}\right)\right)^{p_1^{s}\log^{s}\left(\frac{1}{p_1}\right)}}\right.\right.$$
$$\left.\left.+\sum_{k=2}^{n}\log\left(\frac{\left(P_{k-1}\log\left(\frac{1}{p_k}\right)\right)^{P_{k-1}^{s}\log^{s}\left(\frac{1}{p_k}\right)}}{\left(P_k\log\left(\frac{1}{p_k}\right)\right)^{\left(P_k^{s}\log^{s}\left(\frac{1}{p_k}\right)\right)^{s}}}\right)\right)\right),\quad s\notin\{0,1\}.$$

If Φ_i $(i = 1\text{-}3)$ is positive, then Theorem 11.12 applied for $f = f_s \in \tilde{\Omega}$ and $g = f_q \in \tilde{\Omega}$ yields that there exists $\xi_i \in [a,b]$ such that

$$\xi_i^{s-q} = \frac{\Phi_i(f_s)}{\Phi_i(f_q)}, \qquad i \in \{1,2,3\}.$$

Since the function $\xi_i \mapsto \xi_i^{s-q}$ is invertible for $s \neq q$, we have

$$a \leq \left(\frac{\Phi_i(f_s)}{\Phi_i(f_q)}\right)^{\frac{1}{s-q}} \leq b, \qquad i \in \{1,2,3\},$$

which together with the fact that $\mu_{s,q}\left(\Phi_i, \tilde{\Omega}\right)$ $(i = 1\text{-}3)$ is continuous, symmetric and monotonous (by (11.40)), shows that $\mu_{s,q}\left(\Phi_i, \tilde{\Omega}\right)$ is a mean.

Bibliography

[1] G. Bennett, *Lower bounds for matrices*, Linear Algebra Appl. **82**(1986), 81–98.

[2] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities, Second ed.*, Cambridge University Press, 1952.

[3] J. Jakšetić, Đ Pečarić, J. Pečarić, *Some properties of Zipf-Mandelbrot law and Hurwitz ξ-function*, Math. Inequal. Appl. (2018), 575–584.

[4] C. Jardas, J. Pečarić, R. Roki, N. Sarapa, *On an inequality of Hardy-Littlewood-Pólya and some applications to entropies*, Glas. Mat. Ser. III 32(52)(1997), 201–206.

[5] S. Khalid, J. Pečarić, M. Praljak, *3-convex functions and generalizations of an inequality of Hardy-Littlewood-Pólya*, Glas. Mat. Ser. III **48**(68)(2013), 335–356.

[6] S. Khalid, J. Pečarić, M. Praljak, *On an inequality of I. Perić*, Math. Commun., **19**(2014), 221–242.

[7] S. Khalid, Đ. Pečarić, J. Pečarić, *On Zipf-Mandelbrot entropy and 3-convex functions*, Adv. Oper. Theory (2019), (accepted).

[8] S. Khalid, Đ. Pečarić, J. Pečarić, *On Shannon and Zipf-Mandelbrot entropies and related results*, J. Inequal. Appl. (2019), (accepted).

[9] S. Khalid, J. Pečarić, A. Vukelić, *Refinements of the majorization theorems via Fink identity and related results*, J. Classical. Anal. **7** (2) (2015), 129–154.

[10] M. A. Khan, Đ. Pečarić, J. Pečarić, *Bounds for Shannon and Zipf-Mandelbrot entropies*, Math. Meth Appl Sci. (2017), 1–7.

[11] N. Latif, Đ. Pečarić, J. Pečarić, *Majorization, Csiszar Divergence and Zipf-Mandelbrot law*, J. Inequal. Appl. (2017), 1–15.

[12] M. G. Mann, C. Tsallis, *Nonextensive Entropy: Interdisciplinary Applications*, Oxford University Press, 2004.

[13] M. A. Montemurro, *Beyond the Zipf-Mandelbrot law in quantitative linguistics*, Physica A **300**(2001), 567–578.

[14] J. Pečarić, J. Perić, *Improvements of the Giaccardi and the Petrović inequality and related Stolarsky type means*, An. Univ. Craiova Ser. Mat. Inform. **39**(2012), 65–75.

[15] J. Pečarić, F. Proschan, Y. L. Tong, *Convex Functions, Partial Orderings, and Statistical Applications*, Academic Press Inc., 1992.

List of Authors

Zaid Mohammad Al-Sahwi
Department of Mathematics
University of Saadah
Saadah, Yemen
alsahwizaid@gmail.com

Ana Barbir
Faculty of Civil Engineering,
Architecture And Geodesy
University of Split
Split, Croatia
Ana.Barbir@gradst.hr

Slavica Ivelić Bradanović
Faculty of Civil Engineering,
Architecture And Geodesy
University of Split
Split, Croatia
sivelic@gradst.hr

Saad Ihsan Butt
Department of Mathematics
COMSATS University Islamabad
Lahore Campus
Lahore, Pakistan
saadihsanbutt@gmail.com

László Horváth
Department of Mathematics
University of Pannonia Egyetem
Veszprém, Hungary
lhorvath@almos.uni-pannon.hu

Julije Jakšetić
Faculty of Food Technology
and Biotechnology
Mathematics department
University of Zagreb
Zagreb, Croatia
jjaksetic@pbf.hr

Sadia Khalid
Department of Mathematics
COMSATS University Islamabad
Lahore Campus
Lahore, Pakistan
saadiakhalid176@gmail.com
skhalid@cuilahore.edu.pk

Khuram Ali Khan
Department of Mathematics
University of Sargodha
Sargodha, Pakistan
khuramsms@gmail.com

Muhammad Adil Khan
Department of Mathematics
University of Peshawar
Peshawar, Pakistan
adilswati@gmail.com

Neda Lovričević
Faculty of Civil Engineering,
Architecture and Geodesy
University of Split
Split, Croatia
neda.lovricevic@gradst.hr

Rozarija Mikić
Faculty of Textile Technology
University of Zagreb
Zagreb, Croatia
rozarija.jaksic@ttf.hr

Tasadduq Niaz
Department of Mathematics
University of Sargodha
Sargodha, Pakistan
Department of Mathematics
The University of Lahore
Sargodha Campus
Sargodha, Pakistan
tasadduq_khan@yahoo.com

Đilda Pečarić
Catholic University of Croatia
Zagreb, Croatia
gildapeca@gmail.com,
gpecaric@yahoo.com

Josip Pečarić
Zagreb, Croatia
pecaric@element.hr

Jurica Perić
Faculty of Science
Departments of Mathematics
University of Split
Split, Croatia
jperic@pmfst.hr

Dora Pokaz
Faculty of Civil Engineering
University of Zagreb
Zagreb, Croatia
dora@grad.hr

Mirna Rodić
Faculty of Textile Technology
University of Zagreb
Zagreb, Croatia

Ana Vukelić
Faculty of Food Technology
and Biotechnology
Mathematics department
University of Zagreb
Zagreb, Croatia
avukelic@pbf.hr

January, March 2019