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 $2 \le k \le n$ 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 $l \ge 2$ 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*

$$f\left(\sum_{i=1}^{n} p_i v_i\right) \le C_{dis} = C_{dis}(f,\mathbf{v},\mathbf{p},\lambda) \tag{7.3}$$

$$:= \sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right) f\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}v_{i+j}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) \le \sum_{i=1}^{n} p_i f(v_i)$$

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

$$\hat{I}_f(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right)$$

$$\ge \sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) f\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right) \ge f\left(\frac{\sum\limits_{i=1}^{n}p_i}{\sum\limits_{i=1}^{n}q_i}\right)\sum_{i=1}^{n}q_i. \tag{7.5}$$

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 $2 \le k \le n$ 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*

$$-\sum_{i=1}^{n} q_i \log(q_i)$$

$$\le -\sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j} \right) \log \left(\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j} \right) \le \log \left(\frac{n}{\sum_{i=1}^{n} q_i} \right) \sum_{i=1}^{n} q_i. \tag{7.7}$$

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* $\mathbf{q}$

$$H(\mathbf{q}) \le -\sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j} \right) \log \left(\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j} \right) \le \log(n).$$

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 $2 \le k \le n$ 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*

$$\sum_{i=1}^{n} p_i \log \left(\frac{p_i}{q_i} \right) \tag{7.8}$$

$$\ge \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \log \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}}{\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j}} \right) \ge \log \left(\frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i} \right) \sum_{i=1}^{n} p_i. \tag{7.9}$$

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*

$$D(\mathbf{p}\|\mathbf{q}) \geq \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \log \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}}{\sum_{j=0}^{k-1} \lambda_{j+1} q_{i+j}} \right) \geq 0. \tag{7.10}$$

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 $0 \leq \alpha \leq \beta$, $\alpha, \beta \neq 1$*, and the base of* log *is greater than* 1*, then*

$$D_\alpha(\mathbf{p}, \mathbf{q}) \leq \frac{1}{\beta - 1} \log \left(\sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}} \right)^{\alpha-1}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right)^{\frac{\beta-1}{\alpha-1}} \right. \tag{7.13}$$

$$\leq D_\beta(\mathbf{p},\mathbf{q})$$

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

$$\left(\sum_{i=1}^{n} q_i\left(\frac{p_i}{q_i}\right)^{\alpha}\right)^{\frac{1}{\alpha-1}}$$

$$\le \left(\sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right)^{\frac{\beta-1}{\alpha-1}} \right)^{\frac{1}{\beta-1}}$$

$$\le \left(\sum_{i=1}^{n} p_i \left(\frac{p_i}{q_i}\right)^{\beta-1} \right)^{\frac{1}{\beta-1}} = \left(\sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i}\right)^{\beta} \right)^{\frac{1}{\beta-1}}.$$

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* $2 \le k \le n$ *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 $0 \le \alpha < 1$ *and the base of* log *is greater than* 1, *or* $1 < \alpha$ *and the base of* log *is between* 0 *and* 1, *then*

$$\frac{1}{\sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i}\right)^{\alpha}} \sum_{i=1}^{n} p_i \left(\frac{p_i}{q_i}\right)^{\alpha-1} \log\left(\frac{p_i}{q_i}\right) \le \frac{1}{(\alpha-1)\sum_{i=1}^{n} p_i \left(\frac{p_i}{q_i}\right)^{\alpha-1}} \times$$

$$\times \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1} \right) \log \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) \le D_\alpha(\mathbf{p},\mathbf{q}) \quad (7.15)$$

$$\le \frac{1}{\alpha-1} \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \right) \log \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j}} \right) \le D_1(\mathbf{p},\mathbf{q})$$

If either $0 \le \alpha < 1$ *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 $0 \le \alpha < 1$ 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

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

$$\leq \frac{1}{\alpha-1}\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)$$

$$\leq \frac{1}{\alpha-1}\sum_{i=1}^{n} p_i\log\left(\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right) = \sum_{i=1}^{n} p_i\log\left(\frac{p_i}{q_i}\right) = D_1(\mathbf{p},\mathbf{q})$$

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

$$D_\alpha(\mathbf{p},\mathbf{q}) := \frac{1}{\alpha-1}\log\left(\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right)$$

$$= \frac{1}{(\alpha-1)\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\left(\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right)\log\left(\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right)$$

$$\geq \frac{1}{(\alpha-1)\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\right)\times$$

$$\times\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)$$

$$\frac{1}{(\alpha-1)\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\sum_{i=1}^{n}\left(\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}\right)\log\left(\frac{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}\left(\frac{p_{i+j}}{q_{i+j}}\right)^{\alpha-1}}{\sum\limits_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right)$$

$$\geq \frac{1}{(\alpha-1)\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\log\left(\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right)$$

$$= \frac{1}{\sum\limits_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}}\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1}\log\left(\frac{p_i}{q_i}\right)$$

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 $\mathbf{\frac{1}{n}} := \left(\frac{1}{n},\ldots,\frac{1}{n}\right)$ be the discrete uniform distribution.

Corollary 7.3 *Let $2 \le k \le n$ be integers, and let $\lambda := (\lambda_1,\ldots,\lambda_k)$ and $\mathbf{p} := (p_1,\ldots,p_n)$ 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 $\mathbf{q} = \mathbf{\frac{1}{n}}$, 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* $2 \leq k \leq N$ *be integers, and let* $\lambda := (\lambda_1, \ldots, \lambda_k)$ *be a probability distribution. By applying Corollary 7.3 (c), we have:*

If $0 \leq \alpha < 1$, *and the base of* log *is greater than* 1, *then*

$$H_\alpha(\mathbf{p}) = \frac{1}{1-\alpha} \log \left(\frac{1}{H_{N,q,s}^{\alpha}} \sum_{i=1}^{N} \frac{1}{(i+q)^{\alpha s}} \right)$$

$$\geq \frac{1}{1-\alpha} \sum_{i=1}^{n} \left(\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+j+q)^s H_{N,q,s}} \right) \log \left(\frac{1}{H_{N,q,s}^{\alpha-1}} \frac{\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+j+q)^{\alpha s}}}{\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+q)^s}} \right)$$

$$\geq \frac{s}{H_{N,q,s}} \sum_{i=1}^{N} \frac{\log(i+q)}{(i+q)^s} + \log(H_{N,q,s}) = H(\mathbf{p})$$

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* $N \in \{1,2,\ldots\}$, $q_1,\ q_2 \in [0,\infty[$ *and* $s_1,\ s_2 > 0$, *respectively, let* $2 \leq k \leq N$ *be integers, and let* $\lambda := (\lambda_1, \ldots, \lambda_k)$ *be a probability distribution. By applying Corollary 7.2 (b), we have:*

If the base of log *is greater than* 1, *then*

$$D(\mathbf{p}_1 \| \mathbf{p}_2) = \sum_{i=1}^{N} \frac{1}{(i+q_1)^{s_1} H_{N,q_1,s_1}} \log \left(\frac{(i+q_2)^{s_2} H_{N,q_2,s_2}}{(i+q_1)^{s_1} H_{N,q_1,s_1}} \right)$$

$$\geq \sum_{i=1}^{N} \left(\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+q_1)^{s_1} H_{N,q_1,s_1}} \right) \log \left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+q_1)^{s_1} H_{N,q_1,s_1}}}{\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+q_2)^{s_2} H_{N,q_2,s_2}}} \right) \geq 0. \quad (7.17)$$

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 $\dfrac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} z_{i+j}}{\sum_{v=0}^{k-1} \lambda_{j+1} p_{i+j}} \in [\alpha_1, \alpha_2]$ for $i = 1,\dots m$.

(A_2) For the linear functionals $J_u(\cdot)$ $(u = 3,\dots,6)$, assume that $\dfrac{\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} f(z_{i+j})}{\sum_{j=0}^{k-1} \lambda_{v+1} p_{i+j}} \in [\alpha_1, \alpha_2]$ 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 $\sum_{j=0}^{k-1} \lambda_{j+1} p_{i+j} \neq 0$ for $i = 1,\dots m$ with $\sum_{i=1}^{m} p_i = 1$ and $\sum_{j=1}^{k} \lambda_j = 1$. 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)

$$J_u(f(z)) = \int_{\alpha_1}^{\alpha_2} J_u\Big(G_v(z,r)\Big) \sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} f^{(\sigma+2)}(b_\omega) H_{\sigma\omega}(r)dr$$
$$+ \int_{\alpha_1}^{\alpha_2}\int_{\alpha_1}^{\alpha_2} J_u\Big(G_v(z,r)\Big) G_{H,n-2}(r,\xi)) f^{(n)}(\xi) d\xi dr \quad (7.24)$$

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 $u = 1,\ldots,6$,*

$$J_u\Big(G_{H,n}(z,r)\Big) \geq 0, \quad r \in [\alpha_1, \alpha_2] \quad (7.25)$$

holds, then we have

$$J_u(f(z)) \geq \sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} f^{(\sigma)}(b_\omega) J_u\Big(H_{\sigma\omega}(z)\Big) \quad (7.26)$$

for $u = 1,\ldots,6$.

(b) *If for all $u = 1,\ldots,6$ and $v = 1,\ldots,5$*

$$J_u\Big(G_v(z,r)\Big) \geq 0, \; r \in [\alpha_1, \alpha_2] \quad (7.27)$$

holds, provided that s_ω is odd for each $\omega = 2,3,4,\cdots,t$, then

$$J_u(f(z)) \geq \int_{\alpha_1}^{\alpha_2} J_u\Big(G_v(z,r)\Big) \sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} f^{(\sigma+2)}(b_\omega) H_{\sigma\omega}(r)dr. \quad (7.28)$$

for $u = 1,\ldots,6$.

(c) *If (7.27) holds for all $u = 1,\ldots,6$ and $v = 1,\ldots,5$, 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 $u = 1,\ldots,6$.*

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 $m,k \in \mathbb{N}$ $(2 \leq k \leq m)$, $\lambda_1,\ldots,\lambda_k$ 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 $-\infty < \alpha_1 = b_1 < b_2 \cdots < b_t = \alpha_2 < \infty$, $(t \geq 2)$ such that f is n-convex function. Then the following inequalities hold:

$$\hat{I}_f(\mathbf{p},\mathbf{q}) \geq \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) f\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right) + \left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega} f^{(\sigma)}(b_\omega)\right) \times \left(\sum_{i=1}^{m} q_i H_{\sigma\omega}\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) H_{\sigma\omega}\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)\right). \tag{7.33}$$

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 $m,k \in \mathbb{N}$ $(2 \leq k \leq m)$, $\lambda_1,\ldots,\lambda_k$ be positive probability distributions.*

(a) If $\mathbf{q} := (q_1,\ldots,q_m) \in (0,\infty)^m$ and $(n = even)$, then

$$\sum_{i=1}^{m} q_i \ln q_i \geq \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)\ln\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) + \left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right) \times \left(\sum_{i=1}^{m} q_i H_{\sigma\omega}\left(\frac{1}{q_i}\right) - \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right) H_{\sigma\omega}\left(\frac{1}{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)\right). \tag{7.34}$$

(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*

$$D(\mathbf{q} \parallel \mathbf{p}) \geq \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)\ln\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}\right) +$$

$$\left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right)\times$$
$$\left(\sum_{i=1}^{m} q_i H_{\sigma\omega}\left(\frac{p_i}{q_i}\right)-\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}\right)H_{\sigma\omega}\left(\frac{\sum_{j=0}^{k-1}\lambda_{j+1}p_{i+j}}{\sum_{j=0}^{k-1}\lambda_{j+1}q_{i+j}}\right)\right). \quad (7.37)$$

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 $m \in \{1,2,\dots\}$, $t \geq 0$, $s > 0$, then **Zipf-Mandelbrot entropy** can be given as:

$$Z(H,t,s) = \frac{s}{H_{m,t,s}}\sum_{i=1}^{m}\frac{\ln(i+t)}{(i+t)^s} + \ln(H_{m,t,s}). \quad (7.38)$$

Consider

$$q_i = f(i;m,t,s) = \frac{1}{((i+t)^s H_{m,t,s})}. \quad (7.39)$$

Now we state our results involving entropy introduced by Mandelbrot Law:

Theorem 7.10 *Let* $m,k \in \mathbb{N}$ $(2 \leq k \leq m)$*,* $\lambda_1,\dots,\lambda_k$ *be positive probability distributions and* $\mathbf{q}$ *be as defined in* (7.39) *by Zipf-Mandelbrot law with parameters* $m \in \{1,2,\dots\}$*,* $c \geq 0$*,* $d > 0$*. For* $(n = even)$*, the following holds*

$$\begin{aligned}
H(\mathbf{q}) &= Z(H,c,d)\\
&\leq -\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\frac{\lambda_{j+1}}{((i+j+c)^d H_{m,c,d})}\right)\ln\left(\frac{1}{H_{m,c,d}}\sum_{j=0}^{k-1}\frac{\lambda_{j+1}}{((i+j+t)^s)}\right)-\\
&\left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right)\left(\sum_{i=1}^{m}\frac{1}{((i+c)^d H_{m,c,d})}H_{\sigma\omega}\left(((i+c)^d H_{m,c,d})\right)\right)+\\
&\left(\sum_{\omega=1}^{t}\sum_{\sigma=0}^{s_\omega}\frac{(-1)^\sigma(\sigma-1)!}{(b_\omega)^\sigma}\right)\left(\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1}\frac{\lambda_{j+1}}{((i+j+c)^d H_{m,c,d})}\right)H_{\sigma\omega}\left(\frac{1}{\sum_{j=0}^{k-1}\frac{\lambda_{j+1}}{((i+j+c)^d H_{m,c,d})}}\right)\right).
\end{aligned} \quad (7.40)$$

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 $\sum_{i=1}^{m} q_i = 1$. Moreover using above q_i in Shannon entropy (7.3), we get Mandelbrot entropy(7.38). □

Corollary 7.9 *Let $m,k \in \mathbb{N}$ $(2 \leq k \leq m)$, $\lambda_1,\ldots,\lambda_k$ 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*

$$
\begin{aligned}
D(\mathbf{q} \parallel \mathbf{p}) &= \sum_{i=1}^{m} \frac{1}{(i+c_1)^{d_1} H_{m,c_1,d_1}} \ln\left(\frac{(i+c_2)^{d_2} H_{m,c_2,d_2}}{(i+c_1)^{d_1} H_{m,c_1,d_1}}\right) \\
&\geq \sum_{i=1}^{m}\left(\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+j+c_1)^{d_1} H_{m,c_1,d_1}}\right) \ln\left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+c_1)^{d_1} H_{m,c_1,d_1}}}{\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+c_2)^{d_2} H_{m,c_2,d_2}}}\right) \\
&+ \left(\sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} \frac{(-1)^\sigma (\sigma-1)!}{(b_\omega)^\sigma}\right)\left(\sum_{i=1}^{m} \frac{1}{((i+c_1)^{d_1} H_{m,c_1,d_1})} H_{\sigma\omega}\left(\frac{((i+c_2)_2^d H_{m,c_2,d_2})}{((i+c_1)^d H_{m,c_1,d_1})}\right)\right) \\
&- \left(\sum_{\omega=1}^{t} \sum_{\sigma=0}^{s_\omega} \frac{(-1)^\sigma (\sigma-1)!}{(b_\omega)^\sigma}\right) \times \\
&\left(\sum_{i=1}^{m}\left(\sum_{j=0}^{k-1} \frac{\lambda_{j+1}}{(i+j+c_1)^{d_1} H_{m,c_1,d_1}}\right) H_{\sigma\omega}\left(\frac{\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+c_1)^{d_1} H_{m,c_1,d_1}}}{\sum_{j=0}^{k-1} \lambda_{j+1} \frac{1}{(i+j+c_2)^{d_2} H_{m,c_2,d_2}}}\right)\right)
\end{aligned}
\tag{7.41}
$$

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 $\text{essint}_{\nu}(g)$ are the essential infimum $(\text{essinf}_{\nu}(g))$ and the essential supremum of g, and either of them belong to $\text{essint}_{\nu}(g)$ exactly if g takes this value with positive probability.

(b) It is easy to see that either $\text{essint}_{\nu}(g) = \left\{ \int_Y g d\nu \right\}$ (in this case g is constant ν-a.e.) or $\int_Y g d\nu$ is an inner point of $\text{essint}_{\nu}(g)$.

(c) The interval $\text{essinf}_{\nu}(g)$ 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 $\text{essint}_{\nu}(g)$.*

Proof. It is easy to see that the condition is sufficient for equality in (7.42).

Conversely, if $\text{essint}_{\nu}(g)$ contains only one point, then it is trivial, so we can assume that $m := \int_Y g d\nu$ is an inner point of $\text{essint}_{\nu}(g)$. Let

$$l : \mathbb{R} \to \mathbb{R}, \quad l(t) = f'_+(m)(t-m) + f(m).$$

If f is not affine on $\text{essint}_{\nu}(g)$, then by the convexity of f, there is a point $t_1 \in \text{essint}_{\nu}(g)$ 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. □

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 $f \in F$

$$D_f(R_\alpha, Q^n) = \int\limits_{(q>0)^n} f\left(\frac{\sum_{i=1}^{n}\alpha_i q(\omega_1)\ldots p(\omega_i)\ldots q(\omega_n)}{\prod_{i=1}^{n} q(\omega_i)}\right) dQ^n(\omega_1,\ldots,\omega_n)$$

$$+ f^*(0) R_\alpha\left(\bigotimes_{i=1}^{n} q = 0\right)$$

$$= \int\limits_{(q>0)^n} f\left(\sum_{i=1}^{n}\alpha_i \frac{p(\omega_i)}{q(\omega_i)}\right) dQ^n(\omega_1,\ldots,\omega_n) + f^*(0)P(q=0) \tag{7.58}$$

$$= \int\limits_{(q>0)^n} \prod_{i=1}^{n} q(\omega_i) f\left(\sum_{i=1}^{n}\alpha_i \frac{p(\omega_i)}{q(\omega_i)}\right) d\mu^n(\omega_1,\ldots,\omega_n) + f^*(0)P(q=0).$$

By applying Theorem 7.12, we obtain some refinements of the basic inequality 7.51.

Theorem 7.15 *Suppose that* $\alpha_1,\ldots,\alpha_n$ *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

$$\int\limits_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega) \geq \int\limits_{(q>0)^n} f\left(\sum_{i=1}^{n}\alpha_i \frac{p(\omega_i)}{q(\omega_i)}\right) dQ^n(\omega_1,\ldots,\omega_n) \tag{7.60}$$

$$\geq \int\limits_{(q>0)^n} f\left(\frac{1}{n}\sum_{i=1}^{n} \frac{p(\omega_i)}{q(\omega_i)}\right) dQ^n(\omega_1,\ldots,\omega_n),$$

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

$$I(R_\alpha, Q^n) = \infty P(q=0)$$

$$+ \int\limits_{(q>0)^n} \sum_{i=1}^{n}\left(\alpha_i p(\omega_i) \prod_{\substack{j=1\\ j\neq i}}^{n} q(\omega_j)\right) \ln\left(\sum_{i=1}^{n}\alpha_i \frac{p(\omega_i)}{q(\omega_i)}\right) d\mu^n(\omega_1,\ldots,\omega_n),$$

(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 $p_i = \frac{1}{k}$ $(i = 1,\dots,k)$ in the previous corollary (in this case $T = \{1,\dots,k\}$), we have

$$D_f(P,Q) = -\sum_{i=1}^{k} q_i \ln\left(\frac{1}{kq_i}\right) = \ln(k) + \sum_{i=1}^{k} q_i \ln(q_i)$$

$$\geq -\sum_{(i_1,\dots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\frac{1}{k}\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right) = \ln(k) - \sum_{(i_1,\dots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right)$$

$$\geq -\sum_{(i_1,\dots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\frac{1}{kn}\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right)$$

$$= \ln(kn) - \sum_{(i_1,\dots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right) \geq 0.$$

It can be obtained from this some refinements of the classical upper estimation for the Shannon entropy

$$H(Q) := -\sum_{i=1}^{k} q_i \ln(q_i) \leq \sum_{(i_1,\dots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right)$$

$$\leq -\ln(n) + \sum_{(i_1,\dots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right) \leq \ln(k).$$

(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:

$$I(P,Q) = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right) \geq \sum_{(i_1,\dots,i_n)\in T^n} \left(\sum_{j=1}^{n} \alpha_j p_{i_j} \prod_{\substack{l=1\\ l\neq j}}^{n} q_{i_l}\right) \ln\left(\sum_{j=1}^{n} \alpha_j \frac{p_{i_j}}{q_{i_j}}\right)$$

$$\geq \frac{1}{n} \sum_{(i_1,\dots,i_n)\in T^n} \left(\sum_{j=1}^{n} p_{i_j} \prod_{\substack{l=1\\ l\neq j}}^{n} q_{i_l}\right) \ln\left(\frac{1}{n}\sum_{j=1}^{n}\frac{p_{i_j}}{q_{i_j}}\right) \geq 0. \quad (7.61)$$

(c) The Zipf-Mandelbrot law (see Mandelbrot [46] and Zipf [15]) is a discrete probability distribution depends on three parameters $N \in \{1,2,\dots\}$, $q \in [0,\infty[$ and $s > 0$, and it is defined by

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

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

$$I(P,Q) = \sum_{i=1}^{N} \frac{1}{(i+q_1)^{s_1} H_{N,q_1,s_1}} \log\left(\frac{(i+q_2)^{s_2} H_{N,q_2,s_2}}{(i+q_1)^{s_1} H_{N,q_1,s_1}}\right)$$

$$\ge \sum_{(i_1,\ldots,i_N)\in T^n} \left(\sum_{j=1}^{n} \alpha_j \frac{1}{(i_j+q_1)^{s_1} H_{N,q_1,s_1}} \prod_{\substack{l=1\\ l\ne j}}^{n} \frac{1}{(i_l+q_2)^{s_2} H_{N,q_2,s_2}}\right)$$

$$\times \ln\left(\sum_{j=1}^{n} \alpha_j \frac{(i_j+q_2)^{s_2} H_{N,q_2,s_2}}{(i_j+q_1)^{s_1} H_{N,q_1,s_1}}\right) \ge 0.$$

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).