

RESEARCH

Open Access



Majorization, Csiszár divergence and Zipf-Mandelbrot law

Naveed Latif^{1*} , Đilda Pečarić² and Josip Pečarić³

*Correspondence: naveed707@gmail.com
¹Department of Social Studies, Jubail Industrial College, Jubail, 31961, Kingdom of Saudi Arabia
Full list of author information is available at the end of the article

Abstract

In this paper we show how the Shannon entropy is connected to the theory of majorization. They are both linked to the measure of disorder in a system. However, the theory of majorization usually gives stronger criteria than the entropic inequalities. We give some generalized results for majorization inequality using Csiszár f -divergence. This divergence, applied to some special convex functions, reduces the results for majorization inequality in the form of Shannon entropy and the Kullback-Leibler divergence. We give several applications by using the Zipf-Mandelbrot law.

MSC: 94A15; 94A17; 26A51; 26D15

Keywords: majorization inequality; Csiszár f -divergence; Shannon entropy; Kullback-Leibler divergence; Zipf-Mandelbrot law

1 Introduction and preliminaries

Well over a century ago measures were derived for assessing the distance between two models of probability distributions. Most relevant is Boltzmann's [1] concept of generalized entropy in physics and thermodynamics (see Akaike [2] for a brief review). Shannon [3] employed entropy in his famous treatise on communication theory. Kullback-Leibler [4] derived an information measure that happened to be the negative of Boltzmann's entropy, now referred to as the Kullback-Leibler (K-L) distance. The motivation for the Kullback-Leibler work was to provide a rigorous definition of information in relation to Fisher's sufficient statistics. The K-L distance has also been called the K-L discrepancy, divergence, information and number. These terms are synonyms; we use the term 'distance' in the material to follow.

A fundamental result related to the notion of the Shannon entropy is the following inequality (see [5]):

$$\sum_{i=1}^n p_i \log \frac{1}{p_i} \leq \sum_{i=1}^n p_i \log \frac{1}{q_i}, \quad (1)$$

for all positive real numbers p_i and q_i with

$$\sum_{i=1}^n p_i = \sum_{i=1}^n q_i. \quad (2)$$

Here, ‘log’ denotes the logarithmic function taken to a fixed base $b > 1$. Equality holds in (1) if $q_i = p_i$ for all i . For details, see [6], p.635-650. This result, sometimes called the fundamental lemma of information theory, has extensive applications (see for example [7]).

Matić *et al.* [5, 8, 9] and [10] continuously worked on Shannon’s inequality and related inequalities in the probability distribution and information science. They studied and discussed in [5, 10] several aspects of Shannon’s inequality in discrete as well as in integral forms, by presenting upper estimates of the difference between its two sides. Applications to the bounds in information theory were also given.

Now we introduce the main mathematical theory explored in the presented work, the theory of majorization. It is a powerful and elegant mathematical tool which can be applied to a wide variety of problems as well as in quantum mechanics. The theory of majorization is closely related to the notions of randomness and disorder. It indeed allows us to compare two probability distributions in order to know which one is more random. Let us now give the most general definition of majorization.

For fixed $n \geq 2$ let

$$\mathbf{x} = (x_1, \dots, x_n), \quad \mathbf{y} = (y_1, \dots, y_n)$$

denote two real n -tuples. Let

$$\begin{aligned} x_{[1]} \geq x_{[2]} \geq \dots \geq x_{[n]}, & \quad y_{[1]} \geq y_{[2]} \geq \dots \geq y_{[n]}, \\ x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}, & \quad y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)} \end{aligned}$$

denote their ordered components.

The following definition is given in [11], p.319.

Majorization: Let $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n)$ be n -tuples of real numbers. Then we say that \mathbf{y} is majorized by \mathbf{x} or that \mathbf{x} majorizes \mathbf{y} , in symbol, $\mathbf{x} \succ \mathbf{y}$, if we have

$$\sum_{i=1}^j y_{[i]} \leq \sum_{i=1}^j x_{[i]}, \tag{3}$$

for $j = 1, 2, \dots, n - 1$, and

$$\sum_{i=1}^n x_{[i]} = \sum_{i=1}^n y_{[i]}. \tag{4}$$

Note that (3) is equivalent to

$$\sum_{i=n-j+1}^n y_{(i)} \leq \sum_{i=n-j+1}^n x_{(i)},$$

for $j = 1, 2, \dots, n - 1$.

The following theorem, called the classical majorization theorem, is given in the monograph by Marshall *et al.* [12], p.11 (see also [11], p.320):

Theorem 1 (Classical majorization theorem) *Let $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n)$ be two real n -tuples such that $x_i, y_i \in J \subset \mathbb{R}$ for $i = 1, \dots, n$. Then \mathbf{x} majorizes \mathbf{y} if and only if for every continuous convex function $f : J \rightarrow \mathbb{R}$, the following inequality holds:*

$$\sum_{i=1}^n f(y_i) \leq \sum_{i=1}^n f(x_i). \tag{5}$$

The following theorem is a generalization of Theorem 1, known as weighted majorization theorem, and was proved by Fuchs in [13] (see also [11], p.323):

Theorem 2 (Weighted majorization theorem) *Let $\mathbf{x} = (x_1, \dots, x_n), \mathbf{y} = (y_1, \dots, y_n)$ be two decreasing real n -tuples such that $x_i, y_i \in J$ for $i = 1, \dots, n$. Let $\mathbf{w} = (w_1, \dots, w_n)$ be a real n -tuple such that*

$$\sum_{i=1}^j w_i y_i \leq \sum_{i=1}^j w_i x_i, \tag{6}$$

for $j = 1, 2, \dots, n - 1$, and

$$\sum_{i=1}^n w_i y_i = \sum_{i=1}^n w_i x_i. \tag{7}$$

Then, for every continuous convex function $f : J \rightarrow \mathbb{R}$, we have the following inequality:

$$\sum_{i=1}^n w_i f(y_i) \leq \sum_{i=1}^n w_i f(x_i). \tag{8}$$

The following theorem is valid ([14], p.32).

Theorem 3 *Let $f : J \rightarrow \mathbb{R}$ be a continuous convex function on an interval J , \mathbf{w} be a positive n -tuple, and $\mathbf{x}, \mathbf{y} \in J^n$ satisfy*

$$\sum_{i=1}^k w_i y_i \leq \sum_{i=1}^k w_i x_i \quad \text{for } k = 1, \dots, n - 1, \tag{9}$$

and

$$\sum_{i=1}^n w_i y_i = \sum_{i=1}^n w_i x_i. \tag{10}$$

(a) *If \mathbf{y} is a decreasing n -tuple, then*

$$\sum_{i=1}^n w_i f(y_i) \leq \sum_{i=1}^n w_i f(x_i). \tag{11}$$

(b) If \mathbf{x} is an increasing n -tuple, then

$$\sum_{i=1}^n w_i f(x_i) \leq \sum_{i=1}^n w_i f(y_i). \tag{12}$$

If f is strictly convex and $\mathbf{x} \neq \mathbf{y}$, then (11) and (12) are strict.

Matić *et al.* [5, 10] considered a discrete-valued random variable X with finite range $\{x_i\}_{i=1}^r$. Assume $p_i = P\{X = x_i\}$. The b -entropy of X is defined by

$$H_b(X) := \sum_{i=1}^r p_i \log(1/p_i). \tag{13}$$

In [5], they proved that

$$H_b(X) \leq \log r, \tag{14}$$

which shows that the entropy function $H_b(X)$ reaches its maximum value on the discrete uniform probability distribution.

They introduced the idea by giving the general setting of the above inequality by using the classical majorization theorem for the function $f(x) = x \log x$, which is convex and continuous on \mathbb{R}_+ . Suppose X and Y are discrete random variables with finite ranges and probability distributions $\mathbf{p} = \{p_i\}_{i=1}^r$ and $\mathbf{q} = \{q_i\}_{i=1}^r$ ($\sum_{i=1}^r p_i = \sum_{i=1}^r q_i = 1$), such that $\mathbf{p} \succ \mathbf{q}$. Then by the majorization theorem

$$H_b(X) \leq H_b(Y). \tag{15}$$

By substituting $\mathbf{p} \succ (1/r, \dots, 1/r)$ we get (14).

It is generally common to take log with a base of 2 in the introduced notions, but in our investigations this is not essential.

In Section 2, we present our main generalized results obtained from majorization inequality by using Csiszár f -divergence and then obtain corollaries in the form of Shannon entropy and the K-L distance. In Section 3, we give several applications using the Zipf-Mandelbrot law.

2 Csiszár f -divergence for majorization

Csiszár introduced in [15] and then discussed in [16] the following notion.

Definition 1 Let $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a convex function, and let $\mathbf{p} := (p_1, \dots, p_n)$ and $\mathbf{q} := (q_1, \dots, q_n)$ 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 non-negative probability distributions in the f -divergence functional, by defining

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

Horváth *et al.* [17], p.3, considered functionality based on the previous definition.

Definition 2 Let $J \subset \mathbb{R}$ be an interval, and let $f : J \rightarrow \mathbb{R}$ be a function. Let $\mathbf{p} := (p_1, \dots, p_n) \in \mathbb{R}^n$, and $\mathbf{q} := (q_1, \dots, q_n) \in]0, \infty[^n$ be such that

$$\frac{p_i}{q_i} \in J, \quad i = 1, \dots, n. \tag{16}$$

Then we denote

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

Motivated by the ideas in [5] and [10], in this paper we study and discuss the majorization results in the form of divergences and entropies. The following theorem is a generalization of the result given in [5], *i.e.*, (15).

Assume \mathbf{p} and \mathbf{q} to be n -tuples, then we define

$$\frac{\mathbf{p}}{\mathbf{q}} := \left(\frac{p_1}{q_1}, \frac{p_2}{q_2}, \dots, \frac{p_n}{q_n}\right).$$

The following theorem is the connection between Csiszár f -divergence and weighted majorization inequality as one sequence is monotonic.

Theorem 4 Assume $J \subset \mathbb{R}$ to be an interval, $f : J \rightarrow \mathbb{R}$ to be a continuous convex function, p_i, r_i ($i = 1, \dots, n$) to be real numbers and q_i ($i = 1, \dots, n$) to be positive real numbers, such that

$$\sum_{i=1}^k r_i \leq \sum_{i=1}^k p_i, \quad \text{for } k = 1, \dots, n - 1, \tag{17}$$

and

$$\sum_{i=1}^n r_i = \sum_{i=1}^n p_i, \tag{18}$$

with $\frac{p_i}{q_i}, \frac{r_i}{q_i} \in J$ ($i = 1, \dots, n$).

(a) If $\frac{\mathbf{p}}{\mathbf{q}}$ is decreasing, then

$$\hat{I}_f(\mathbf{r}, \mathbf{q}) \leq \hat{I}_f(\mathbf{p}, \mathbf{q}). \tag{19}$$

(b) If $\frac{\mathbf{p}}{\mathbf{q}}$ is increasing, then

$$\hat{I}_f(\mathbf{r}, \mathbf{q}) \geq \hat{I}_f(\mathbf{p}, \mathbf{q}). \tag{20}$$

If f is a continuous concave function, then the reverse inequalities hold in (19) and (20).

Proof (a): We use Theorem 3(a) with substitutions $x_i := \frac{p_i}{q_i}$, $y_i := \frac{r_i}{q_i}$, $w_i := q_i$ and $q_i > 0$ ($i = 1, \dots, n$). Then we get (19).

We can prove part (b) with similar substitutions in Theorem 3(b). □

Theorem 5 *Assume $J \subset \mathbb{R}$ to be an interval, $g : J \rightarrow \mathbb{R}$ to be a function, such that $x \rightarrow xg(x)$ ($x \in J$) to be a continuous convex function, p_i and r_i ($i = 1, \dots, n$) to be real numbers and q_i ($i = 1, \dots, n$) to be positive real numbers satisfying (17) and (18) with*

$$\frac{p_i}{q_i}, \frac{r_i}{q_i} \in J \quad (i = 1, \dots, n).$$

(a) *If $\frac{\mathbf{p}}{\mathbf{q}}$ is decreasing, then*

$$\hat{I}_g(\mathbf{r}, \mathbf{q}) := \sum_{i=1}^n r_i g\left(\frac{r_i}{q_i}\right) \leq \hat{I}_g(\mathbf{p}, \mathbf{q}). \tag{21}$$

(b) *If $\frac{\mathbf{p}}{\mathbf{q}}$ is increasing, then*

$$\hat{I}_g(\mathbf{r}, \mathbf{q}) \geq \hat{I}_g(\mathbf{p}, \mathbf{q}). \tag{22}$$

If $xg(x)$ is a continuous concave function, then the reverse inequalities hold in (21) and (22).

Proof (a): We use Theorem 3(a) with substitutions $x_i = \frac{p_i}{q_i}$, $y_i = \frac{r_i}{q_i}$, $w_i = q_i$ as $q_i > 0$ ($i = 1, \dots, n$), and $f(x) := xg(x)$. Then we get (21).

We can prove part (b) with similar substitutions in Theorem 3(b) for $f(x) := xg(x)$. □

The theory of majorization and the notion of entropic measure of disorder are closely related. Based on this fact, the aim of this paper is to look for majorization relations with the connection to entropic inequalities. This was interesting to do for two main reasons. The first one is the fact that the majorization relations are usually stronger than the entropic inequalities, in the sense that they imply these entropic inequalities, but the converse is not true. The second reason is the fact that, when we dispose of majorization relations between two different quantum states, we know that we can transform one of the states into the other using some unitary transformation. The concept of entropy alone would not allow us to prove such a property.

The Shannon entropy was introduced in the field of classical information. There are two ways of viewing the Shannon entropy. Suppose we have a random variable X , and we learn its value. In one point of view, the Shannon entropy quantifies the amount of information as regards the value of X (after measurement). In another point of view, the Shannon entropy tells us the amount of uncertainty about the variable of X before we learn its value (before measurement).

We mention two special cases of the previous result.

The first case corresponds to the entropy of a discrete probability distribution.

Definition 3 The Shannon entropy of a positive probability distribution $\mathbf{p} := (p_1, \dots, p_n)$ is defined by

$$H(\mathbf{p}) := - \sum_{i=1}^n p_i \log p_i. \tag{23}$$

Note that there is no problem with the definition in the case of a zero probability, since

$$\lim_{x \rightarrow 0} x \log x = 0. \tag{24}$$

Corollary 1 Assume p_i, r_i and q_i ($i = 1, \dots, n$) to be positive real numbers satisfying (17) and (18) with

$$\frac{p_i}{q_i}, \frac{r_i}{q_i} \in J \quad (i = 1, \dots, n).$$

- (a) If $\frac{\mathbf{p}}{\mathbf{q}}$ is a decreasing n -tuple and the base of \log is greater than 1, then the following estimates for the Shannon entropy of \mathbf{q} hold:

$$\sum_{i=1}^n q_i \log \left(\frac{r_i}{q_i} \right) \geq H(\mathbf{q}). \tag{25}$$

If the base of \log is in between 0 and 1, then the reverse inequality holds in (25).

- (b) If $\frac{\mathbf{p}}{\mathbf{q}}$ is an increasing n -tuple and the base of \log is greater than 1, then the following estimates for the Shannon entropy of \mathbf{q} hold:

$$H(\mathbf{q}) \leq \sum_{i=1}^n q_i \log \left(\frac{p_i}{q_i} \right). \tag{26}$$

If the base of \log is in between 0 and 1, then the reverse inequality holds in (26).

Proof (a): Substitute $f(x) := \log x$ and $p_i = 1$ ($i = 1, \dots, n$) in Theorem 4(a). Then we get (25).

We can prove the part (b) with similar substitutions for $r_i = 1$ ($i = 1, \dots, n$). □

Corollary 2 Assume p_i and r_i ($i = 1, \dots, n$) to be positive real numbers satisfying (17) and (18).

- (a) If \mathbf{r} is a decreasing n -tuple and the base of \log is greater than 1, then for the connection between the Shannon entropies of \mathbf{p} and \mathbf{r}

$$H(\mathbf{r}) \geq H(\mathbf{p}). \tag{27}$$

If the base of \log is in between 0 and 1, then the reverse inequality holds in (27).

- (b) If \mathbf{p} is an increasing n -tuple and the base of \log is greater than 1, then for the connection between Shannon entropies of \mathbf{p} and \mathbf{r}

$$H(\mathbf{r}) \leq H(\mathbf{p}). \tag{28}$$

If the base of \log is in between 0 and 1, then the reverse inequality holds in (28).

Proof (a): Substitute $g(x) := \log x$ and $q_i = 1$ ($i = 1, \dots, n$) in Theorem 5(a). Then we get (27). We can prove part (b) with similar substitutions. \square

The second case corresponds to the relative entropy or the K-L distance between two probability distributions.

Definition 4 The K-L distance between the positive probability distributions $\mathbf{p} := (p_1, \dots, p_n)$ and $\mathbf{q} := (q_1, \dots, q_n)$ is defined by

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

Corollary 3 Assume $J \subset \mathbb{R}$ to be an interval, and p_i, r_i and q_i ($i = 1, \dots, n$) to be positive real numbers satisfying (17) and (18) with

$$\frac{p_i}{q_i}, \frac{r_i}{q_i} \in J \quad (i = 1, \dots, n).$$

(a) If $\frac{\mathbf{r}}{\mathbf{q}}$ is a decreasing n -tuple and the base of \log is greater than 1, then

$$\sum_{i=1}^n q_i \log\left(\frac{r_i}{q_i}\right) \geq \sum_{i=1}^n q_i \log\left(\frac{p_i}{q_i}\right). \tag{29}$$

If the base of \log is in between 0 and 1, then the reverse inequality holds in (29).

(b) If $\frac{\mathbf{p}}{\mathbf{q}}$ is an increasing n -tuple and the base of \log is greater than 1, then

$$\sum_{i=1}^n q_i \log\left(\frac{r_i}{q_i}\right) \leq \sum_{i=1}^n q_i \log\left(\frac{p_i}{q_i}\right). \tag{30}$$

If the base of \log is in between 0 and 1, then the reverse inequality holds in (30).

Proof (a): Substitute $f(x) := \log x$ in Theorem 4(a). Then we get (29). We can prove part (b) with substitution $f(x) := \log x$ in Theorem 4(b). \square

Corollary 4 Let $J \subset \mathbb{R}$ be an interval and assume p_i, r_i and q_i ($i = 1, \dots, n$) be positive real numbers satisfying (17) and (18) with

$$\frac{p_i}{q_i}, \frac{r_i}{q_i} \in J \quad (i = 1, \dots, n).$$

(a) If $\frac{\mathbf{r}}{\mathbf{q}}$ is a decreasing n -tuple and the base of \log is greater than 1, then the following comparison inequality between K-L distance of (\mathbf{r}, \mathbf{q}) and (\mathbf{p}, \mathbf{q}) holds:

$$L(\mathbf{r}, \mathbf{q}) := \sum_{i=1}^n r_i \log\left(\frac{r_i}{q_i}\right) \leq L(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right). \tag{31}$$

If the base of \log is in between 0 and 1, then the reverse inequality holds in (31).

- (b) If $\frac{\mathbf{p}}{\mathbf{q}}$ is an increasing n -tuple and the base of \log is greater than 1, then the following comparison inequality between K-L distance of (\mathbf{r}, \mathbf{q}) and (\mathbf{p}, \mathbf{q}) holds:

$$\sum_{i=1}^n r_i \log\left(\frac{r_i}{q_i}\right) \geq \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right). \tag{32}$$

If the base of \log is in between 0 and 1, then the reverse inequality holds in (32).

Proof (a): Substitute $g(x) := \log x$ in Theorem 5(a). Then we get (31).

We can prove part (b) with substitution $g(x) := \log x$ in Theorem 5(b). □

Remark 1 We give the above results when one sequence is monotone by using Theorem 3, but we can give all the above results when both sequences are monotone by using the weighted majorization theorem, Theorem 2, for $w_i > 0$ ($i = 1, \dots, n$).

3 Applications to the Zipf-Mandelbrot entropy

The term Zipfian distribution refers to a distribution of probabilities of occurrence that follows Zipf’s law. Zipf’s law is an experimental law, not a theoretical one; *i.e.* it describes an occurrence rather than predicting it from some kind of theory: the observation that, in many natural and man-made phenomena, the probability of occurrence of many random items starts high and tapers off. Thus, a few occur very often while many others occur rarely. The formal definition of this law is $\mathbf{P}_n = 1/n^a$, where \mathbf{P}_n is the frequency of occurrence of the n th ranked item and a is close to 1.

Converted to language, this means that the rank of a word (in terms of its frequency) is approximately inversely proportional to its actual frequency, and so produces a hyperbolic distribution. To put Zipf’s law in another way (see [18, 19]): $fr = C$, where r = the rank of a word, f = the frequency of occurrence of that word, and C = a constant (the value of which depends on the subject under consideration). Essentially this shows an inverse proportional relationship between a word’s frequency and its frequency rank. Zipf called this curve the ‘standard curve’. Texts from natural languages do not, of course, behave with such absolute mathematical precision. They can not, because, for one thing, any curve representing empirical data from large texts will be a stepped graph, since many non-high-frequency words will share the same frequency. But the overall consensus is that texts match the standard curve significantly well. Li [20] writes ‘this distribution, also called Zipf’s law, has been checked for accuracy for the standard corpus of the present-day English [Kućera and Francis] with very good results.’ See Miller [21] for a concise summary of the match between actual data and the standard curve.

Zipf also studied the relationship between the frequency of occurrence of a word and its length. In *The Psycho-Biology of Language*, he stated that ‘it seems reasonably clear that shorter words are distinctly more favored in language than longer words.’

Apart from the use of this law in information science and linguistics, Zipf’s law is used in economics. This distribution in economics is known as Pareto’s law, which analyzes the distribution of the wealthiest members of the community [22], p.125. These two laws are the same in the mathematical sense, but they are applied in different contexts [23], p.294. The same type of distribution that we have in Zipf’s and Pareto’s law, also known as the power law, can also be found in other scientific disciplines, such as physics, biology, earth and planetary sciences, computer science, demography and the social sciences [24].

Benoit Mandelbrot in [25] gave a generalization of Zipf’s law, now known as the Zipf-Mandelbrot law, which gave an improvement in the account for the low-rank words in a corpus where $k < 100$ [26]:

$$f(k) = \frac{C}{(k + q)^s},$$

when $q = 0$, we get Zipf’s law.

For $n \in \mathbb{N}$, $q \geq 0$, $s > 0$, $k \in \{1, 2, \dots, n\}$, in a more clear form, the Zipf-Mandelbrot law (probability mass function) is defined with

$$f(k, n, q, s) := \frac{1/(k + q)^s}{H_{n,q,s}}, \tag{33}$$

where

$$H_{n,q,s} := \sum_{i=1}^n \frac{1}{(i + q)^s}, \tag{34}$$

$n \in \mathbb{N}$, $q \geq 0$, $s > 0$, $k \in \{1, 2, \dots, n\}$.

Application of the Zipf-Mandelbrot law can also be found in linguistics [26], information sciences [19, 23] and ecological field studies [27].

In probability theory and statistics, the cumulative distribution function (CDF) of a real-valued random variable X , or just distribution function of X , evaluated at x , is the probability that X will take a value less than or equal to x and we often denote by CDF the following ratio:

$$\text{CDF} := \frac{H_{k,t,s}}{H_{n,t,s}}. \tag{35}$$

The cumulative distribution function is an important application of majorization.

In the case of a continuous distribution, it gives the area under the probability distribution functions, also used to specify the distribution of multivariable random variables.

There are various applications of CDF. For example, in learning to rank, the CDF arises naturally as a probability measure over inequality events of the type $\{X \leq x\}$. The joint CDF lends itself to problems that are easily described in terms of inequality events in which statistical dependence relationships also exist among events. Examples of this type of problem include web search and document retrieval [28–31], predicting rating of movies [32] or predicting multiplayer game outcomes with a team structure [33]. In contrast to the canonical problems of classification or regression, in learning to rank we are required to learn some mapping from inputs to inter-dependent output variables, so that we may wish to model both stochastic orderings of variable states and statistical dependence relationships between variables.

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

Application 1 Assume \mathbf{p} and \mathbf{r} to be the Zipf-Mandelbrot laws with parameters $n \in \{1, 2, \dots\}$, $t_1, t_2 \geq 0$ and $s_1, s_2 > 0$, respectively, satisfying

$$\frac{H_{k,t_2,s_2}}{H_{n,t_2,s_2}} \leq \frac{H_{k,t_1,s_1}}{H_{n,t_1,s_1}}, \quad k = 1, \dots, n - 1, \tag{36}$$

and also let $q_i > 0$ ($i = 1, 2, \dots, n$).

(a) If $\frac{(i+t_2)^{s_2}}{(i+1+t_2)^{s_2}} \leq \frac{q_{i+1}}{q_i}$ ($i = 1, \dots, n$) and the base of log is greater than 1, then

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \log \left(\frac{1}{q_i (i+t_2)^{s_2} H_{n,t_2,s_2}} \right) \\ & \leq \sum_{i=1}^n \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log \left(\frac{1}{q_i (i+t_1)^{s_1} H_{n,t_1,s_1}} \right). \end{aligned} \tag{37}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (37).

(b) If $\frac{(i+t_1)^{s_1}}{(i+1+t_1)^{s_1}} \geq \frac{q_{i+1}}{q_i}$ ($i = 1, \dots, n$) and the base of log is greater than 1, then

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \log \left(\frac{1}{q_i (i+t_2)^{s_2} H_{n,t_2,s_2}} \right) \\ & \geq \sum_{i=1}^n \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log \left(\frac{1}{q_i (i+t_1)^{s_1} H_{n,t_1,s_1}} \right). \end{aligned} \tag{38}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (38).

Proof (a) Assume $p_i := \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $r_i := \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, then

$$\sum_{i=1}^k p_i := \sum_{i=1}^k \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} = \frac{1}{H_{n,t_1,s_1}} \sum_{i=1}^k \frac{1}{(i+t_1)^{s_1}} = \frac{H_{k,t_1,s_1}}{H_{n,t_1,s_1}}, \quad k = 1, \dots, n - 1.$$

Similarly, $\sum_{i=1}^k r_i := \frac{H_{k,t_2,s_2}}{H_{n,t_2,s_2}}, k = 1, \dots, n - 1$.

This implies that

$$\sum_{i=1}^k r_i \leq \sum_{i=1}^k p_i \Leftrightarrow \frac{H_{k,t_2,s_2}}{H_{n,t_2,s_2}} \leq \frac{H_{k,t_1,s_1}}{H_{n,t_1,s_1}}, \quad k = 1, \dots, n - 1.$$

We can easily check that $\frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ is decreasing over $i = 1, \dots, n$ and similarly r_i too.

Now, we investigate the behavior of $\frac{\mathbf{r}}{\mathbf{q}}$ for $q_i > 0$ ($i = 1, 2, \dots, n$); take

$$\begin{aligned} \frac{r_i}{q_i} &= \frac{1}{q_i (i+t_2)^{s_2} H_{n,t_2,s_2}} \quad \text{and} \quad \frac{r_{i+1}}{q_{i+1}} = \frac{1}{q_{i+1} (i+1+t_2)^{s_2} H_{n,t_2,s_2}}, \\ \frac{r_{i+1}}{q_{i+1}} - \frac{r_i}{q_i} &= \frac{1}{H_{n,t_2,s_2}} \left[\frac{1}{q_{i+1} (i+1+t_2)^{s_2}} - \frac{1}{q_i (i+t_2)^{s_2}} \right] \leq 0 \\ &\Leftrightarrow \frac{(i+t_2)^{s_2}}{(i+1+t_2)^{s_2}} \leq \frac{q_{i+1}}{q_i}, \end{aligned}$$

which shows that $\frac{r}{q}$ is decreasing. So all the assumptions of Corollary 4(a) are true. Then by using (31) we get (37).

(b) If we switch the role of r_i into p_i , then by using (32) in Corollary 4(b) we get (38). \square

The following application is a special case of the above result.

Application 2 Assume \mathbf{p} and \mathbf{r} to be the Zipf-Mandelbrot laws with parameters $n \in \{1, 2, \dots\}$, $t_1, t_2 \geq 0$ and $s_1, s_2 > 0$, respectively, satisfying (36).

If the base of log is greater than 1, then

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}} \log\left(\frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}\right) \\ & \leq \sum_{i=1}^n \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log\left(\frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}\right). \end{aligned} \tag{39}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (39).

Proof Substitute $q_i := 1$ ($i = 1, 2, \dots, n$) in (37); we get (39). \square

Application 3 Assume \mathbf{p} and \mathbf{r} to be the Zipf-Mandelbrot laws with parameters $n \in \{1, 2, \dots\}$, $t_1, t_2 \geq 0$ and $s_1, s_2 > 0$, respectively, satisfying (36) and also let $q_i > 0$ ($i = 1, 2, \dots, n$).

(a) If $\frac{(i+t_2)^{s_2}}{(i+1+t_2)^{s_2}} \leq \frac{q_{i+1}}{q_i}$ ($i = 1, \dots, n$) and the base of log is greater than 1, then

$$\sum_{i=1}^n q_i \log\left(\frac{1}{q_i(i+t_2)^{s_2} H_{n,t_2,s_2}}\right) \geq \sum_{i=1}^n q_i \log\left(\frac{1}{q_i(i+t_1)^{s_1} H_{n,t_1,s_1}}\right). \tag{40}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (40).

(b) If $\frac{(i+t_1)^{s_1}}{(i+1+t_1)^{s_1}} \geq \frac{q_{i+1}}{q_i}$ ($i = 1, \dots, n$) and the base of log is greater than 1, then

$$\sum_{i=1}^n q_i \log\left(\frac{1}{q_i(i+t_2)^{s_2} H_{n,t_2,s_2}}\right) \leq \sum_{i=1}^n q_i \log\left(\frac{1}{q_i(i+t_1)^{s_1} H_{n,t_1,s_1}}\right). \tag{41}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (41).

Proof We can prove by a similar method as given in Application 1 with substitutions $p_i := \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $r_i := \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$ in Corollary 3 instead of Corollary 4, to get the required results. \square

The following result is a special case of Application 3.

Application 4 Assume \mathbf{p} and \mathbf{r} to be the Zipf-Mandelbrot laws with parameters $n \in \{1, 2, \dots\}$, $t_1, t_2 \geq 0$ and $s_1, s_2 > 0$, respectively, satisfying (36). If the base of log is greater than 1, then

$$\sum_{i=1}^n \log\left(\frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}\right) \geq \sum_{i=1}^n \log\left(\frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}\right). \tag{42}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (42).

Proof Substitute $q_i := 1$ ($i = 1, 2, \dots, n$) in (40); we get (42). □

Application 5 Assume \mathbf{p} and \mathbf{r} to be the Zipf-Mandelbrot laws with parameters $n \in \{1, 2, \dots\}$, $t_1, t_2 \geq 0$ and $s_1, s_2 > 0$, respectively, satisfying (36), and also let $q_i > 0$ ($i = 1, 2, \dots, n$).

(a) If $\frac{(i+t_2)^{s_2}}{(i+1+t_2)^{s_2}} \leq \frac{q_{i+1}}{q_i}$ ($i = 1, \dots, n$) and the base of log is greater than 1, then

$$\sum_{i=1}^n q_i \log \left(\frac{1}{q_i(i+t_2)^{s_2} H_{n,t_2,s_2}} \right) \geq H(\mathbf{q}). \tag{43}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (43).

(b) If $\frac{(i+t_1)^{s_1}}{(i+1+t_1)^{s_1}} \geq \frac{q_{i+1}}{q_i}$ ($i = 1, \dots, n$) and the base of log is greater than 1, then

$$H(\mathbf{q}) \leq \sum_{i=1}^n q_i \log \left(\frac{1}{q_i(i+t_1)^{s_1} H_{n,t_1,s_1}} \right). \tag{44}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (44).

Proof (a) We can prove (43), by a similar method to that given in Application 1, with substitutions $p_i := 1$ and $r_i := \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$, in Corollary 1(a) instead of Corollary 4(a).

(b) For this part, switch the role of \mathbf{p} and \mathbf{r} in part (a), like $p_i := \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$ and $r_i := 1$ ($i = 1, 2, \dots, n$), and applying Corollary 1(b) instead of Corollary 4(b), we get (44). □

At the end, in the following application, we use three of the Zipf-Mandelbrot laws for different parameters.

Application 6 Assume \mathbf{p}, \mathbf{q} and \mathbf{r} to be the Zipf-Mandelbrot laws with parameters $n \in \{1, 2, \dots\}$, $t_1, t_2, t_3 \geq 0$ and $s_1, s_2, s_3 > 0$, respectively, satisfying (36).

(a) If $\frac{(i+1+t_2)^{s_2}}{(i+1+t_3)^{s_3}} \leq \frac{(i+t_2)^{s_2}}{(i+t_3)^{s_3}}$ ($i = 1, \dots, n$) and the base of log is greater than 1, then

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{(i+t_3)^{s_3} H_{n,t_3,s_3}} \log \left(\frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_3)^{s_3} H_{n,t_3,s_3}} \right) \\ & \leq \sum_{i=1}^n \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log \left(\frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \right). \end{aligned} \tag{45}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (45).

(b) If $\frac{(i+1+t_2)^{s_2}}{(i+1+t_3)^{s_3}} \geq \frac{(i+t_2)^{s_2}}{(i+t_3)^{s_3}}$ ($i = 1, \dots, n$) and the base of log is greater than 1, then

$$\begin{aligned} & \sum_{i=1}^n \frac{1}{(i+t_3)^{s_3} H_{n,t_3,s_3}} \log \left(\frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_3)^{s_3} H_{n,t_3,s_3}} \right) \\ & \geq \sum_{i=1}^n \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \log \left(\frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_1)^{s_1} H_{n,t_1,s_1}} \right). \end{aligned} \tag{46}$$

If the base of log is in between 0 and 1, then the reverse inequality holds in (46).

Proof (a) Let $p_i := \frac{1}{(i+t_1)^{s_1} H_{n,t_1,s_1}}$, $q_i := \frac{1}{(i+t_2)^{s_2} H_{n,t_2,s_2}}$ and $r_i := \frac{1}{(i+t_3)^{s_3} H_{n,t_3,s_3}}$. Here p_i, q_i and r_i are decreasing over $i = 1, \dots, n$. Now, we investigate the behavior of $\frac{r}{q}$.

Take

$$\begin{aligned} \frac{r_i}{q_i} &= \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_3)^{s_3} H_{n,t_3,s_3}} \quad \text{and} \quad \frac{r_{i+1}}{q_{i+1}} = \frac{(i+1+t_2)^{s_2} H_{n,t_2,s_2}}{(i+1+t_3)^{s_3} H_{n,t_3,s_3}}, \\ \frac{r_{i+1}}{q_{i+1}} - \frac{r_i}{q_i} &= \frac{(i+1+t_2)^{s_2} H_{n,t_2,s_2}}{(i+1+t_3)^{s_3} H_{n,t_3,s_3}} - \frac{(i+t_2)^{s_2} H_{n,t_2,s_2}}{(i+t_3)^{s_3} H_{n,t_3,s_3}}, \\ \frac{r_{i+1}}{q_{i+1}} - \frac{r_i}{q_i} &= \frac{H_{n,t_2,s_2}}{H_{n,t_3,s_3}} \left[\frac{(i+1+t_2)^{s_2}}{(i+1+t_3)^{s_3}} - \frac{(i+t_2)^{s_2}}{(i+t_3)^{s_3}} \right]; \end{aligned}$$

the R.H.S. is non-positive by using the assumption, which shows that $\frac{r}{q}$ is decreasing, therefore using Corollary 4(a) we get (45).

(b) If we replace $\frac{r}{q}$ with $\frac{p}{q}$ in part (a) and use Corollary 4(b), we get (46). □

4 Conclusions

In this paper we show how the Shannon entropy is connected to the theory of majorization. They are both linked to the measure of disorder in a system. However, the theory of majorization usually gives stronger criteria than the entropic inequalities. The theory of majorization and the notion of entropic measure of disorder are closely related. Based on this fact, the aim of this paper is to look for majorization relations with entropic inequalities. We give some generalized results for Csiszár f -divergence of majorization inequality. We mention two special cases of these generalized results; the first case corresponds to the entropy of a discrete probability distribution, and the second case corresponds to the relative entropy or the Kullback-Leibler distance between two probability distributions. The cumulative distribution function (CDF) is an important application of majorization. We give several applications by using the Zipf-Mandelbrot law with (CDF).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

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

Author details

¹Department of Social Studies, Jubail Industrial College, Jubail, 31961, Kingdom of Saudi Arabia. ²Catholic University of Croatia, Ilica 242, Zagreb, 10000, Croatia. ³Faculty of Textile Technology Zagreb, University of Zagreb, Prilaz Baruna Filipovića 30, Zagreb, 10000, Croatia.

Publisher's Note

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

Received: 24 April 2017 Accepted: 9 August 2017 Published online: 24 August 2017

References

1. Boltzmann, L: Ueber die Beziehung zwischen dem Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respective den Sätzen über das Warmegleichgewicht. *Wien. Ber. Nat.wiss. Kunst* **76**, 373-435 (1877)
2. Akaike, H: Prediction and Entropy. Springer, New York (1985)
3. Shannon, CE: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379-423 and 623-656 (1948)
4. Kullback, S, Leibler, RA: On information and sufficiency. *Ann. Math. Stat.* **22**(1), 79-86 (1951)
5. Matić, M, Pearce, CEM, Pečarić, J: Shannon's and related inequalities in information theory. In: Rassias, TM (ed.) *Survey on Classical Inequalities*, pp. 127-164. Kluwer Academic, Norwell (2000)
6. Mitrinović, DS, Pečarić, J, Fink, AM: *Classical and New Inequalities in Analysis*. Kluwer Academic, Dordrecht (1993)
7. McEliece, RJ: *The Theory of Information and Coding*. Addison-Wesley, Reading (1977)

8. Matić, M, Pearce, CEM, Pečarić, J: Improvements of some bounds on entropy measures in information theory. *Math. Inequal. Appl.* **1**, 295-304 (1998)
9. Matić, M, Pearce, CEM, Pečarić, J: On an inequality for the entropy of a probability distribution. *Acta Math. Hung.* **85**, 345-349 (1999)
10. Matić, M, Pearce, CEM, Pečarić, J: Some refinements of Shannon's inequalities. *ANZIAM J.* **43**, 493-511 (2002)
11. Pečarić, J, Proschan, F, Tong, YL: *Convex Functions, Partial Orderings and Statistical Applications*. Academic Press, New York (1992)
12. Marshall, AW, Olkin, I, Arnold, BC: *Inequalities: Theory of Majorization and Its Applications*, 2nd edn. Springer Series in Statistics. Springer, New York (2011)
13. Fuchs, L: A new proof of an inequality of Hardy-Littlewood-Polya. *Math. Tidsskr.* 53-54 (1947)
14. Niculescu, CP, Persson, LE: *Convex Functions and Their Applications, a Contemporary Approach*. CMS Books in Mathematics, vol. 23. Springer, New York (2006)
15. Csiszár, I: Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hung.* **2**, 299-318 (1967)
16. Csiszár, I: Information measure: a critical survey. In: *Trans. 7th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Processes and 8th European Meeting of Statist.*, B, pp. 73-86. Academia, Prague (1978)
17. Horváth, L, Pečarić, Đ, Pečarić, J: Estimations of f- and Rényi divergences by using a cyclic refinement of the Jensen's inequality. *Bull. Malays. Math. Sci. Soc.* (2017). doi:10.1007/s40840-017-0526-4
18. Adil Khan, M, Pečarić, Đ, Pečarić, J: Bounds for Shannon and Zipf-Mandelbrot entropies. *Math. Methods Appl. Sci.* to appear
19. Silagadze, ZK: Citations and the Zipf-Mandelbrot law. *Complex Syst.* **11**, 487-499 (1997)
20. Li, W: Random texts exhibits Zipf's-law-like word frequency distribution. *IEEE Trans. Inf. Theory* **38**(6), 1842-1845 (1992)
21. Miller, GA: *Language and Communication*. McGraw-Hill, New York (1951)
22. Diodato, V: *Dictionary of Bibliometrics*. Haworth Press, New York (1994)
23. Egghe, L, Rousseau, R: *Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science*. Elsevier, New York (1990)
24. Newmann, MEJ: Power laws, Pareto distributions and Zipf's law. arXiv:cond-mat/0412004
25. Mandelbrot, B: Information theory and psycholinguistics: a theory of words frequencies. In: Lazafeld, P, Henry, N (eds) *Reading in Mathematical Social Science*. MIT Press, Cambridge (1966)
26. Montemurro, MA: Beyond the Zipf-Mandelbrot law in quantitative linguistics (2001). arXiv:cond-mat/0104066v2
27. Mouillot, D, Lepretre, A: Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity. *Environ. Monit. Assess.* **63**(2), 279-295 (2000)
28. Burgas, CJC, Shaked, T, Renshaw, E, Lazier, A, Deeds, M, Hamilton, N, Hallender, G: Learning to rank using gradient descent. In: *Proceedings of the Twenty-Second International Conference on Machine Learning (ICML)*, pp. 89-96 (2005)
29. Cao, Z, Qin, T, Liu, TY, Tsai, MF, Li, H: Learning to rank: from pair-wise approach to listwise approach. In: *Proceedings of the Twenty-Second International Conference on Machine Learning (ICML)*, pp. 129-136 (2007)
30. Joachims, T: A support vector method for multivariable performance measures. In: *Proceedings of the Twenty-Second International Conference on Machine Learning (ICML)*, pp. 377-384 (2005)
31. Xia, F, Liu, TY, Wang, J, Zhang, W, Li, H: Listwise approach to learning to rank - theory and algorithms. In: *Proceedings of the Twenty-Second International Conference on Machine Learning (ICML)*, pp. 1192-1199 (2008)
32. Rennie, J, Srebro, N: Fast maximum margin matrix factorization for collaborative prediction. In: *Proceedings of the Twenty-Second International Conference on Machine Learning (ICML)*, pp. 713-719 (2005)
33. Herbrich, R, Minka, TP, Graepel, T: TrueSkill™: a Bayesian skill rating system. *Adv. Neural Inf. Process. Syst.* **19**, 569-576 (2007)

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

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

Submit your next manuscript at ► springeropen.com
