CrossMark

Zipf–Mandelbrot law, f-divergences and the Jensen-type interpolating inequalities

Neda Lovričević[1*] , Đilda Pečarić[2] and Josip Pečarić[3,4]

*Correspondence:
neda.lovricevic@gradst.hr
[1] Faculty of Civil Engineering,
Architecture and Geodesy,
University of Split, Split, Croatia
Full list of author information is
available at the end of the article

Abstract

Motivated by the method of interpolating inequalities that makes use of the improved Jensen-type inequalities, in this paper we integrate this approach with the well known Zipf–Mandelbrot law applied to various types of f-divergences and distances, such are Kullback–Leibler divergence, Hellinger distance, Bhattacharyya distance (via coefficient), χ^2-divergence, total variation distance and triangular discrimination. Addressing these applications, we firstly deduce general results of the type for the Csiszár divergence functional from which the listed divergences originate. When presenting the analyzed inequalities for the Zipf–Mandelbrot law, we accentuate its special form, the Zipf law with its specific role in linguistics. We introduce this aspect through the Zipfian word distribution associated to the English and Russian languages, using the obtained bounds for the Kullback–Leibler divergence.

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

Keywords: Jensen inequality; Zipf and Zipf–Mandelbrot law; Csiszár divergence functional; f-divergences; Kullback–Leibler divergence

1 Introduction

Let us start with the notion of f-divergences which measure the distance between two probability distributions by making an average value, which is weighted by a specific function, of the odds ratio given by two probability distributions. Among the existing f-divergences introduced in the process of finding the adequate distance between two probability distributions, let us point out the Csiszár f-divergence [1, 2], some special cases of which are the Kullback–Leibler divergence (see [3, 4]), the Hellinger distance, the Bhattacharyya distance, the total variation distance, the triangular discrimination (see [5, 6]). The notion of 'distance' mainly appears as somewhat stronger than 'divergence' since it suggests the properties of symmetry and the triangle inequality. Considering a great number of fields in which probability theory cooperates, it is no wonder that divergences between probability distributions have many specific applications in a variety of those fields.

Jensen's inequality, on the other hand, with its numerous refinements, variants and improvements is often called 'a king inequality', obviously not without a reason. Here we try to integrate one of such results concerning Jensen's inequality in order to obtain new estimates for mentioned divergences (which deal with the convex functions for the most

part). It is well known that the Jensen inequality

$$f\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right) \le \frac{1}{P_n}\sum_{i=1}^{n}p_i f(x_i) \tag{1}$$

holds for a convex function $f \colon I \to \mathbb{R}$, $I \subseteq \mathbb{R}$, an n-tuple $\mathbf{x} = (x_1,\dots,x_n) \in I^n$, $n \ge 2$ and nonnegative n-tuple $\mathbf{p} = (p_1,\dots,p_n)$, such that $P_n = \sum_{i=1}^{n}p_i > 0$.

Here we cite a result of Pečarić [7, p. 717] who investigated the method of interpolating inequalities which have reverse inequalities of Aczél type. Using Jensen's inequality and its reverse he proved the main and the following deduced result, which holds for a convex function f defined on an interval $I \subset \mathbb{R}$:

$$\begin{aligned}
\min_{1 \le i \le n} \{p_i\} &\left[\sum_{i=1}^{n} f(x_i) - nf\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right) \right] \\
&\le \sum_{i=1}^{n}p_i f(x_i) - P_n f\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right) \\
&\le \max_{1 \le i \le n}\{p_i\}\left[\sum_{i=1}^{n}f(x_i) - nf\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right) \right],
\end{aligned} \tag{2}$$

where $\mathbf{x} = (x_1,\dots,x_n) \in I^n$, $n \ge 2$ and a nonnegative n-tuple $\mathbf{p} = (p_1,\dots,p_n)$ is such that $P_n = \sum_{i=1}^{n}p_i > 0$.

In recent investigations of relation (2) and its numerous consequences, it appeared as a fruitful field for many significant results. We accentuate those which deal with this relation in view of superadditivity and monotonicity of the Jensen-type functionals, in [8, 9] or [10], obtained via [11] and suitably summarized in the monograph [12]. In the following part we are going to make use of relation (2) while presenting certain bounds for a selected spectrum of f-divergences that originate from the Csiszár divergence functional.

All of the results thus obtained concerning f-divergences are going to be observed here in the context of the Zipf–Mandelbrot law and then specified for the Zipf law.

George Kingsley Zipf (1902–1950) was a linguist after whom one of the most common laws in probability and statistics was named. Today this experimental law for the discrete probability distribution frequently is used in information science, bibliometrics, linguistics, social sciences, economy (where it is known as the Pareto law), as well as in physics, biology, computer science *etc*. Thus the term 'Zipfian distribution' is used to describe various types of distributions of the probability occurrences which approximately follow the mathematical form of the Zipf law. It was in the first place established with the frequency of the words in a text in view and as such it revealed a hyperbolic relation. As is *e.g.* explained in [13], if words of a language are sorted in the order of decreasing frequency of usage, a word's frequency f is inversely proportional to its rank r, or sequence number in the list and the product of these is a constant: $r \cdot f = C$ ('A few occur very often while many others occur rarely.'). Benoit Mandelbrot, a mathematician very well known for his contribution in fractal theory, generalized the Zipf law in 1966 [14, 15] according to his field of investigation and gave its improvement for the count of the low-rank words [16]. It is also used in information sciences for the purpose of indexing [17, 18], in ecological

field studies [19] and it plays its role in art when determining the aesthetics criteria in music [20]. The Zipf–Mandelbrot law is a discrete probability distribution and is defined by the following probability mass function:

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

(3)

where

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

(4)

is a generalization of a harmonic number and $N \in \{1, 2, \dots\}$, $s > 0$ and $t \in [0, \infty)$ are parameters.

For finite N and for $t = 0$ the Zipf–Mandelbrot law is simply called the Zipf law. (In particular, if we observe the infinite N and $t = 0$ we actually have the Zeta distribution.)

According to the expressions above, the probability mass function referring to the Zipf law is

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

(5)

The rest of the paper is organized as follows. In Section 1 we define the Csiszár functional and various f-divergences for which we give in Section 3 the results based on relation (2). These are further examined in Section 4 in the light of the Zipf–Mandelbrot law and the Zipf law. For the latter we give in Section 5 a specific application in linguistics, concerning the Kullback–Leibler divergence.

2 Preliminaries

The previously mentioned f-divergences were studied independently by several matematicians. Here we focus on the Csiszár f-divergences. Csiszár [1, 2] introduced the f-divergence functional as

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

(6)

where $\mathbf{p} = (p_1, \dots, p_n)$ and $\mathbf{q} = (q_1, \dots, q_n)$ are probability distributions, that is, $p_i, q_i \in [0, 1]$, for $i = 1, \dots, n$ with $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$ and $f : [0, \infty) \to [0, \infty)$ is a convex function, the so-called 'distance function' on the set of all probability distributions.

As in [1], we interpret the undefined expressions by

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

The definition of the f-divergence functional (6) can be generalized for a function $f : I \to \mathbb{R}$, $I \subseteq \mathbb{R}$, where $\frac{p_i}{q_i} \in I$, for every $i = 1, \dots, n$. Since we are going to observe this wider class of functions as well, the corresponding functional (6) will be denoted by $\tilde{D}_f(\mathbf{p}, \mathbf{q})$ (see also [21]).

The general aspect of the Csiszár divergence functional (6) can be interpreted as a series of well-known entropies, divergencies and distances, for special choices of the kernel f. In the sequel we present some of the most frequent among them.

Entropies quantify the diversity, uncertainty and randomness of a system. The concept of the Rényi entropy was introduced by [22] and has been of a great importance in statistics, ecology, theoretical computer science *etc.*

The Rényi entropy of order α of $\mathbf{p}$ is defined as

$$H_\alpha(p) := \frac{1}{1-\alpha} \log\left(\sum_{i=1}^n p_i^\alpha\right), \tag{7}$$

where $\alpha \geq 0$, $\alpha \neq 1$ and $\mathbf{p} = (p_1, \ldots, p_n)$ is a probability distribution. Among special cases of the Rényi entropy (*e.g.* the Hartley or max-entropy, min-entropy, and the collision entropy), the Rényi entropy tends to the Shannon entropy (see [23]) for the limiting value of $\alpha \to 1$. The Shannon entropy (which is sometimes called information divergence) is thus defined as

$$H(\mathbf{p}) := -\sum_{i=1}^n p_i \log p_i. \tag{8}$$

Besides the absolute entropies, one can also observe the relative entropies, as did Rényi when he introduced a special form of f-divergence. The Rényi divergence of order α, $\alpha \geq 0$, $\alpha \neq 1$ for the probability distributions $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ is defined as

$$D_\alpha(\mathbf{p}, \mathbf{q}) := \frac{1}{\alpha - 1} \log\left(\sum_{i=1}^n q_i \left(\frac{p_i}{q_i}\right)^\alpha\right). \tag{9}$$

A relation similar to the one between the Rényi entropy and the Shannon entropy holds in the case of the Rényi divergence and the Kullback–Leibler divergence (see [24]) for the probability distributions $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$. As $\alpha \to 1$, the Rényi divergence tends to the Kullback–Leibler divergence. The latter is sometimes called the relative entropy and is defined by

$$\mathrm{KL}(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^n p_i \log\left(\frac{p_i}{q_i}\right). \tag{10}$$

Remark 1 Although it is common to take the logarithm function with the base 2, it will not be essential in the sequel. Moreover, we are going to analyze the results including the logarithm function for different (positive) bases, namely, for those greater than 1 as well as for those that are less than 1.

Among various divergences and considering the properties of symmetry and triangular inequality which some of them possess, we can also define certain distances between two probability distributions.

Thus the Hellinger distance between the probability distributions $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ is defined by

$$h(\mathbf{p}, \mathbf{q}) := \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2}. \tag{11}$$

The Hellinger distance is a metric and is often used in its squared form, *i.e.* as $h^2(\mathbf{p}, \mathbf{q}) := \frac{1}{2} \sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2$.

Among the values of the order α for the Rényi divergence, some have wider application than others. The value 1 is already determined by continuity in α since it cannot be calculated directly by (9) and another interesting example is the order 1/2. This order makes the Rényi divergence symmetric in its arguments. In this context, it is interesting to see how the Rényi divergence, although not itself a metric, relates to the Hellinger distance:

$$D_{1/2}(\mathbf{p}, \mathbf{q}) = -2 \log \left(1 - \frac{h^2(\mathbf{p}, \mathbf{q})}{2} \right). \tag{12}$$

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

$$B(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} \sqrt{p_i q_i}, \tag{13}$$

whereas the Bhattacharyya distance is defined as $D_B(\mathbf{p}, \mathbf{q}) := -\log B(\mathbf{p}, \mathbf{q})$. The relation between the Bhattacharyya coefficient and the Hellinger distance is

$$h^2(\mathbf{p}, \mathbf{q}) = 1 - B(\mathbf{p}, \mathbf{q}).$$

In order to conclude this overview, let us remind the reader that the χ^2 divergence is defined as

$$\chi^2(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}, \tag{14}$$

the total variation distance or statistical distance is given by

$$V(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} |p_i - q_i| \tag{15}$$

and the definition of the triangular discrimination reads as follows:

$$\Delta(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i}. \tag{16}$$

More detailed analyses of the mentioned divergences as well as their wider spectrum one can find *e.g.* in [5, 6, 24].

3 Basic relations for *f*-divergences

In order to deduce the relations from relation (2) for the f-divergences described in the introductory part, we start with the general result for bounds obtained for the Csiszár functional (6) observed under more general conditions as $\tilde{D}_f(\mathbf{p}, \mathbf{q})$.

Theorem 1 *Let $I \subseteq \mathbb{R}$ be an interval. Suppose $\mathbf{p} = (p_1, \ldots, p_n)$ is an n-tuple of real numbers with $P_n = \sum_{i=1}^{n} p_i$ and $\mathbf{q} = (q_1, \ldots, q_n)$ is an n-tuple of nonnegative real numbers with $Q_n = \sum_{i=1}^{n} q_i$, such that $\frac{p_i}{q_i} \in I$, $i = 1, \ldots, n$.*

If $f : I \to \mathbb{R}$ is a convex function, then

$$
\min\{q_i\} \left[\sum_{i=1}^{n} f\left(\frac{p_i}{q_i}\right) - nf\left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i}\right) \right] + Q_n f\left(\frac{P_n}{Q_n}\right)
$$

$$
\leq \tilde{D}_f(\mathbf{p}, \mathbf{q}) \leq \max\{q_i\} \left[\sum_{i=1}^{n} f\left(\frac{p_i}{q_i}\right) - nf\left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i}\right) \right] + Q_n f\left(\frac{P_n}{Q_n}\right). \tag{17}
$$

If f is a concave function, then the inequality signs are reversed.

If $t \mapsto tf(t)$ is a convex function, then

$$
\min\{q_i\} \left[\sum_{i=1}^{n} \frac{p_i}{q_i} f\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^{n} \frac{p_i}{q_i} f\left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i}\right) \right] + P_n f\left(\frac{P_n}{Q_n}\right)
$$

$$
\leq \tilde{D}_{\mathrm{id}\cdot f}(\mathbf{p}, \mathbf{q}) \leq \max\{q_i\} \left[\sum_{i=1}^{n} \frac{p_i}{q_i} f\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^{n} \frac{p_i}{q_i} f\left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i}\right) \right] + P_n f\left(\frac{P_n}{Q_n}\right), \tag{18}
$$

where $\tilde{D}_{\mathrm{id}\cdot f}(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{n} p_i f(\frac{p_i}{q_i})$.

If $t \mapsto tf(t)$ is a concave function, then the inequality signs are reversed.

Proof If we observe a convex function f and replace p_i by q_i as well as x_i by $\frac{p_i}{q_i}$ in relation (2), we get (17).

If we then observe the function $t \mapsto tf(t)$ as a convex function and replace p_i by q_i and x_i by $\frac{p_i}{q_i}$ we get (18). □

The following corollary precedes the related result for the Kullback–Leibler divergence (10). Recall that (10) can be interpreted as a special case of the functional (6).

Corollary 1 *Let $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ be n-tuples of nonnegative real numbers with $P_n = \sum_{i=1}^{n} p_i$ and $Q_n = \sum_{i=1}^{n} q_i$. Then*

$$
\min\{q_i\} \left[\sum_{i=1}^{n} \frac{p_i}{q_i} \log \frac{p_i}{q_i} - \sum_{i=1}^{n} \frac{p_i}{q_i} \log\left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i}\right) \right] + P_n \log \frac{P_n}{Q_n}
$$

$$
\leq \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}
$$

$$
\leq \max\{q_i\} \left[\sum_{i=1}^{n} \frac{p_i}{q_i} \log \frac{p_i}{q_i} - \sum_{i=1}^{n} \frac{p_i}{q_i} \log\left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i}\right) \right] + P_n \log \frac{P_n}{Q_n}, \tag{19}
$$

where the logarithm base is greater than 1.

If the logarithm base is less than 1, *then the inequality signs are reversed.*

Proof It follows from Theorem 1 as a special case of inequalities (18), for the function $t \mapsto t \log t$, which is convex when the logarithm base is greater than 1 (and concave when the base is less than 1). □

If we additionally specify the n-tuples **p** and **q** as in the sequel, we provide the bounds for the Kullback–Leibler divergence.

Remark 2 If we observe $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ as probability distributions, we may write

$$
\min\{q_i\}\left[\sum_{i=1}^{n} \frac{p_i}{q_i} \log \frac{p_i}{q_i} - \sum_{i=1}^{n} \frac{p_i}{q_i} \log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{q_i}\right)\right]
$$

$$
\leq \mathrm{KL}(\mathbf{p},\mathbf{q}) \leq \max\{q_i\}\left[\sum_{i=1}^{n} \frac{p_i}{q_i} \log \frac{p_i}{q_i} - \sum_{i=1}^{n} \frac{p_i}{q_i} \log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{q_i}\right)\right], \quad (20)
$$

where the logarithm base is greater than 1.

If the logarithm base is less than 1, then the inequality signs are reversed.

In other words, we obtained the corresponding bounds for the *Kullback–Leibler divergence* (10).

Remark 3 The Kullback–Leibler divergence is sometimes used in its reversed form $\mathrm{KL}(\mathbf{q},\mathbf{p})$. A similar type of bounds can be obtained when observing the reversed Kullback–Leibler divergence making use of the kernel function $f(t) = -\log t$, its convexity and concavity related to the observed logarithm base (greater than 1 or less than 1, respectively), and following the analogous procedure described in Corollary 1 and Remark 2.

It is natural to observe in a similar fashion the other divergences (distances) described in Section 1: the Hellinger distance, the Bhattacharyya coefficient, the chi-square divergence, the total variation distance and the triangular discrimination.

Corollary 2 *Let* $\mathbf{p} = (p_1, \ldots, p_n)$ *and* $\mathbf{q} = (q_1, \ldots, q_n)$ *be n-tuples of nonnegative real numbers with* $P_n = \sum_{i=1}^{n} p_i$ *and* $Q_n = \sum_{i=1}^{n} q_i$. *Then*

$$
\frac{1}{2}\min\{q_i\}\left[\sum_{i=1}^{n}\left(1 - \sqrt{\frac{p_i}{q_i}}\right)^2 - n\left(1 - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{q_i}}\right)^2\right] + Q_n\left(1 - \sqrt{\frac{P_n}{Q_n}}\right)^2
$$

$$
\leq \frac{1}{2}\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2
$$

$$
\leq \frac{1}{2}\max\{q_i\}\left[\sum_{i=1}^{n}\left(1 - \sqrt{\frac{p_i}{q_i}}\right)^2 - n\left(1 - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{q_i}}\right)^2\right] + Q_n\left(1 - \sqrt{\frac{P_n}{Q_n}}\right)^2. \quad (21)
$$

Proof It follows from Theorem 1 as a special case of inequalities (17), for the convex function $t \to \frac{1}{2}(\sqrt{t} - 1)^2$. □

Remark 4 If we observe $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ as probability distributions, we may write

$$\frac{1}{2}\min\{q_i\}\left[\sum_{i=1}^{n}\left(1 - \sqrt{\frac{p_i}{q_i}}\right)^2 - n\left(1 - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{q_i}}\right)^2\right]$$

$$\leq h^2(\mathbf{p}, \mathbf{q})$$

$$\leq \frac{1}{2}\max\{q_i\}\left[\sum_{i=1}^{n}\left(1 - \sqrt{\frac{p_i}{q_i}}\right)^2 - n\left(1 - \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{p_i}{q_i}}\right)^2\right]. \tag{22}$$

In other words, we obtained the corresponding bounds for the *(squared) Hellinger distance* $h^2(\mathbf{p}, \mathbf{q})$.

Corollary 3 *Let* $\mathbf{p} = (p_1, \ldots, p_n)$ *and* $\mathbf{q} = (q_1, \ldots, q_n)$ *be n-tuples of nonnegative real numbers with* $P_n = \sum_{i=1}^{n} p_i$ *and* $Q_n = \sum_{i=1}^{n} q_i$. *Then*

$$\min\{q_i\}\left[\sqrt{n\sum_{i=1}^{n}\frac{p_i}{q_i}} - \sum_{i=1}^{n}\sqrt{\frac{p_i}{q_i}}\right] - \sqrt{P_n Q_n}$$

$$\leq -\sum_{i=1}^{n}\sqrt{p_i q_i}$$

$$\leq \max\{q_i\}\left[\sqrt{n\sum_{i=1}^{n}\frac{p_i}{q_i}} - \sum_{i=1}^{n}\sqrt{\frac{p_i}{q_i}}\right] - \sqrt{P_n Q_n}. \tag{23}$$

Proof It follows from Theorem 1 as a special case of inequalities (17), for the convex function $f(t) = -\sqrt{t}$. □

Remark 5 If we observe $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ as probability distributions and adopt by the definition (13) that $B(\mathbf{p}, \mathbf{q}) = \sqrt{p_i q_i}$, we may write

$$\min\{q_i\}\left[\sqrt{n\sum_{i=1}^{n}\frac{p_i}{q_i}} - \sum_{i=1}^{n}\sqrt{\frac{p_i}{q_i}}\right] - 1$$

$$\leq -B(\mathbf{p}, \mathbf{q})$$

$$\leq \max\{q_i\} \leq \left[\sqrt{n\sum_{i=1}^{n}\frac{p_i}{q_i}} - \sum_{i=1}^{n}\sqrt{\frac{p_i}{q_i}}\right] - 1, \tag{24}$$

or

$$1 - \min\{q_i\}\left[\sqrt{n\sum_{i=1}^{n}\frac{p_i}{q_i}} - \sum_{i=1}^{n}\sqrt{\frac{p_i}{q_i}}\right]$$

$$\geq B(\mathbf{p}, \mathbf{q}) \geq 1 - \max\{q_i\}\left[\sqrt{n\sum_{i=1}^{n}\frac{p_i}{q_i}} - \sum_{i=1}^{n}\sqrt{\frac{p_i}{q_i}}\right]. \tag{25}$$

In other words, we obtained the corresponding bounds for the *Bhattacharyya coefficient* $B(\mathbf{p}, \mathbf{q})$.

Corollary 4 *Let* $\mathbf{p} = (p_1, \ldots, p_n)$ *be an n-tuple of real numbers and* $\mathbf{q} = (q_1, \ldots, q_n)$ *an n-tuple of nonnegative real numbers with* $P_n = \sum_{i=1}^{n} p_i$ *and* $Q_n = \sum_{i=1}^{n} q_i$. *Then*

$$
\min\{q_i\} \left[\sum_{i=1}^{n} \left(\frac{p_i}{q_i} - 1 \right)^2 - n \left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1 \right)^2 \right] + \frac{(P_n - Q_n)^2}{Q_n}
$$

$$
\leq \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}
$$

$$
\leq \max\{q_i\} \left[\sum_{i=1}^{n} \left(\frac{p_i}{q_i} - 1 \right)^2 - n \left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1 \right)^2 \right] + \frac{(P_n - Q_n)^2}{Q_n}. \tag{26}
$$

Proof It follows from Theorem 1 as a special case of inequalities (17), for the convex function $f(t) = (t - 1)^2$. □

Remark 6 If we observe $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ as probability distributions, we may write

$$
\min\{q_i\} \left[\sum_{i=1}^{n} \left(\frac{p_i}{q_i} - 1 \right)^2 - n \left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1 \right)^2 \right]
$$

$$
\leq \chi^2(\mathbf{p}, \mathbf{q})
$$

$$
\leq \max\{q_i\} \left[\sum_{i=1}^{n} \left(\frac{p_i}{q_i} - 1 \right)^2 - n \left(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1 \right)^2 \right]. \tag{27}
$$

In other words, we obtained the corresponding bounds for the *chi-square divergence* $\chi^2(\mathbf{p}, \mathbf{q})$.

Corollary 5 *Let* $\mathbf{p} = (p_1, \ldots, p_n)$ *be an n-tuple of real numbers and* $\mathbf{q} = (q_1, \ldots, q_n)$ *an n-tuple of nonnegative real numbers with* $P_n = \sum_{i=1}^{n} p_i$ *and* $Q_n = \sum_{i=1}^{n} q_i$. *Then*

$$
\min\{q_i\} \left[\sum_{i=1}^{n} \left| \frac{p_i}{q_i} - 1 \right| - n \left| \frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1 \right| \right] + |P_n - Q_n|
$$

$$
\leq \sum_{i=1}^{n} |p_i - q_i|
$$

$$
\leq \max\{q_i\} \left[\sum_{i=1}^{n} \left| \frac{p_i}{q_i} - 1 \right| - n \left| \frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1 \right| \right] + |P_n - Q_n|. \tag{28}
$$

Proof It follows from Theorem 1 as a special case of inequalities (17), for the convex function $f(t) = |t - 1|$. □

Remark 7 If we observe $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ as probability distributions, we may write

$$\min\{q_i\} \left[\sum_{i=1}^{n} \left| \frac{p_i}{q_i} - 1 \right| - n \left| \frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1 \right| \right]$$

$$\leq V(\mathbf{p}, \mathbf{q})$$

$$\leq \max\{q_i\} \left[\sum_{i=1}^{n} \left| \frac{p_i}{q_i} - 1 \right| - n \left| \frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1 \right| \right]. \tag{29}$$

In other words, we obtained the corresponding bounds for the *total variation distance* $V(\mathbf{p}, \mathbf{q})$.

Corollary 6 *Let* $\mathbf{p} = (p_1, \ldots, p_n)$ *be an n-tuple of real numbers and* $\mathbf{q} = (q_1, \ldots, q_n)$ *an n-tuple of nonnegative real numbers with* $P_n = \sum_{i=1}^{n} p_i$ *and* $Q_n = \sum_{i=1}^{n} q_i$. *Then*

$$\min\{q_i\} \left[\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i(p_i + q_i)} - n \frac{(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1)^2}{\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} + 1} \right] + \frac{(P_n - Q_n)^2}{P_n + Q_n}$$

$$\leq \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{p_i + q_i}$$

$$\leq \max\{q_i\} \left[\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i(p_i + q_i)} - n \frac{(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1)^2}{\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} + 1} \right] + \frac{(P_n - Q_n)^2}{P_n + Q_n}. \tag{30}$$

Proof It follows from Theorem 1 as a special case of inequalities (17), for the convex function $f(t) = \frac{(t-1)^2}{t+1}$. □

Remark 8 If we observe $\mathbf{p} = (p_1, \ldots, p_n)$ and $\mathbf{q} = (q_1, \ldots, q_n)$ as probability distributions, we may write

$$\min\{q_i\} \left[\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i(p_i + q_i)} - n \frac{(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1)^2}{\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} + 1} \right]$$

$$\leq \Delta(\mathbf{p}, \mathbf{q})$$

$$\leq \max\{q_i\} \left[\sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i(p_i + q_i)} - n \frac{(\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} - 1)^2}{\frac{1}{n} \sum_{i=1}^{n} \frac{p_i}{q_i} + 1} \right]. \tag{31}$$

In other words, we obtained the corresponding bounds for the *triangular discrimination* $\Delta(\mathbf{p}, \mathbf{q})$.

4 On *f*-divergences for the Zipf–Mandelbrot law

In this section we are going to derive the results from the previous section for the Zipf–Mandelbrot law (3). Namely, if we put $q_i = f(i; N, s, t)$ in (3) as its probability mass function, we can observe obtained results in the light of the Zipf–Mandelbrot law.

For this purpose we present the general results concerning the Csiszár functional $\tilde{D}_f(\mathbf{p}, \mathbf{q})$ for the Zipf–Mandelbrot law. If we define $\mathbf{q}$ via (3) as a Zipf–Mandelbrot law

N-tuple, the definition (6) of the Csiszár functional becomes

$$\tilde{D}_f(i,N,s_2,t_2,\mathbf{p}) = \sum_{i=1}^{N} \frac{1}{(i+t_2)^{s_2}H_{N,s_2,t_2}} f\left(p_i(i+t_2)^{s_2}H_{N,s_2,t_2}\right), \tag{32}$$

where $f: I \to \mathbb{R}$, $I \subseteq \mathbb{R}$, and the parameters $N \in \mathbb{N}$, $s_2 > 0$, $t_2 \geq 0$ are such that $p_i(i + t_2)^{s_2}H_{N,s_2,t_2} \in I$, $i = 1,\ldots,N$.

The Csiszár functional (6) assumes the following form when $\mathbf{p}$ and $\mathbf{q}$ are both defined as Zipf–Mandelbrot law N-tuples:

$$\tilde{D}_f(i,N,s_1,s_2,t_1,t_2) = \sum_{i=1}^{N} \frac{1}{(i+t_2)^{s_2}H_{N,s_2,t_2}} f\left(\frac{(i+t_2)^{s_2}H_{N,s_2,t_2}}{(i+t_1)^{s_1}H_{N,s_1,t_1}} \right), \tag{33}$$

where $f: I \to \mathbb{R}$, $I \subseteq \mathbb{R}$, and $N \in \mathbb{N}$, $s_1,s_2 > 0$, $t_1,t_2 \geq 0$ are such that $\frac{(i+t_2)^{s_2}H_{N,s_2,t_2}}{(i+t_1)^{s_1}H_{N,s_1,t_1}} \in I$, $i = 1,\ldots,N$.

Finally, both $\mathbf{p}$ and $\mathbf{q}$ N-tuples may be defined via the Zipf law (5) and thus the Csiszár functional (6) assumes the form

$$\tilde{D}_f(i,N,s_1,s_2) = \sum_{i=1}^{N} \frac{1}{i^{s_2}H_{N,s_2}} f\left(i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}} \right). \tag{34}$$

Our next step is providing the corresponding forms of Theorem 1 which are suitable for further applications. Thus we start with the Csiszár functional $\tilde{D}_f(i,N,s_2,t_2,\mathbf{p})$, which implies single Zipf–Mandelbrot laws q_i, for $i = 1,\ldots,N$.

Corollary 7 *Let $\mathbf{p} = (p_1,\ldots,p_N)$ be an N-tuple of real numbers with $P_N = \sum_{i=1}^{N} p_i$. Suppose $I \subseteq \mathbb{R}$ is an interval, $N \in \mathbb{N}$ and $s_2 > 0$, $t_2 \geq 0$ are such that $p_i(i+t_2)^{s_2}H_{N,s_2,t_2} \in I$, $i = 1,\ldots,N$.*

If $f: I \to \mathbb{R}$ is a convex function, then

$$\frac{1}{(N+t_2)^{s_2}H_{N,s_2,t_2}}\Lambda_1 + P_Nf(P_N) \leq \tilde{D}_f(i,N,s_2,t_2,\mathbf{p})$$

$$\leq \frac{1}{(1+t_2)^{s_2}H_{N,s_2,t_2}}\Lambda_1 + P_Nf(P_N), \tag{35}$$

where

$$\Lambda_1 = \sum_{i=1}^{N} f\left(p_i(i+t_2)^{s_2}H_{N,s_2,t_2}\right) - Nf\left(\frac{1}{N}\sum_{i=1}^{N} p_i(i+t_2)^{s_2}H_{N,s_2,t_2} \right).$$

If f is a concave function, then the inequality signs are reversed.

If $t \mapsto tf(t)$ is a convex function, then

$$\frac{1}{(N+t_2)^{s_2}H_{N,s_2,t_2}}\bar{\Lambda}_1 + P_Nf(P_N) \leq \tilde{D}_{\mathrm{id}\cdot f}(i,N,s_2,t_2,\mathbf{p})$$

$$\leq \frac{1}{(1+t_2)^{s_2}H_{N,s_2,t_2}}\bar{\Lambda}_1 + P_Nf(P_N), \tag{36}$$

where $\tilde{D}_{\mathrm{id}\cdot f}(i,N,s_2,t_2,\mathbf{p}) := \sum_{i=1}^{N} p_i f\big(p_i(i+t_2)^{s_2} H_{N,s_2,t_2}\big)$ and

$$\bar{\Lambda}_1 = \sum_{i=1}^{N} p_i(i+t_2)^{s_2} H_{N,s_2,t_2} f\big(p_i(i+t_2)^{s_2} H_{N,s_2,t_2}\big)$$
$$- \sum_{i=1}^{N} p_i(i+t_2)^{s_2} H_{N,s_2,t_2} f\left(\frac{1}{N}\sum_{i=1}^{N} p_i(i+t_2)^{s_2} H_{N,s_2,t_2}\right).$$

If $t \mapsto tf(t)$ is a concave function, then the inequality signs are reversed.

Proof It leans on the proof of Theorem 1 with its described substitutions, where we insert for q_i the expression $\frac{1}{(i+t_2)^{s_2} H_{N,s_2,t_2}}, i = 1,\dots,N$, which defines the Zipf–Mandelbrot law (3), with $Q_N = 1$. Since the minimal value for q_i is $\min\{q_i\} = \frac{1}{(N+t_2)^{s_2} H_{N,s_2,t_2}}$ and its maximal value is $\max\{q_i\} = \frac{1}{(1+t_2)^{s_2} H_{N,s_2,t_2}}$, inequalities (35) and (36) follow for the convex functions f and $t \mapsto tf(t)$, respectively. They change their signs in the case of concavity as a consequence of the Jensen inequality implicitly included. □

If we have both $\mathbf{p}$ and $\mathbf{q}$ defined via the Zipf–Mandelbrot law, then the following corollary plays a role.

Corollary 8 *Let $I \subseteq \mathbb{R}$ be an interval and suppose $N \in \mathbb{N}$, $s_1, s_2 > 0$, $t_1, t_2 \geq 0$ are such that $\frac{(i+t_2)^{s_2} H_{N,s_2,t_2}}{(i+t_1)^{s_1} H_{N,s_1,t_1}} \in I, i = 1,\dots,N$.*
If $f : I \to \mathbb{R}$ is a convex function, then

$$\frac{1}{(N+t_2)^{s_2} H_{N,s_2,t_2}}\Lambda_2 + f(1) \leq \tilde{D}_f(i,N,s_1,s_2,t_1,t_2) \leq \frac{1}{(1+t_2)^{s_2} H_{N,s_2,t_2}}\Lambda_2 + f(1), \quad (37)$$

where

$$\Lambda_2 = \sum_{i=1}^{N} f\left(\frac{(i+t_2)^{s_2} H_{N,s_2,t_2}}{(i+t_1)^{s_1} H_{N,s_1,t_1}}\right) - Nf\left(\frac{1}{N}\sum_{i=1}^{N}\frac{(i+t_2)^{s_2} H_{N,s_2,t_2}}{(i+t_1)^{s_1} H_{N,s_1,t_1}}\right).$$

If f is a concave function, then the inequality signs are reversed.
If $t \mapsto tf(t)$ is a convex function, then

$$\frac{1}{(N+t_2)^{s_2} H_{N,s_2,t_2}}\bar{\Lambda}_2 + f(1) \leq \tilde{D}_{\mathrm{id}\cdot f}(i,N,s_1,s_2,t_1,t_2) \leq \frac{1}{(1+t_2)^{s_2} H_{N,s_2,t_2}}\bar{\Lambda}_2 + f(1), \quad (38)$$

where $\tilde{D}_{\mathrm{id}\cdot f}(i,N,s_1,s_2,t_1,t_2) := \sum_{i=1}^{N}\frac{1}{(i+t_1)^{s_1} H_{N,s_1,t_1}} f\big(\frac{(i+t_2)^{s_2} H_{N,s_2,t_2}}{(i+t_1)^{s_1} H_{N,s_1,t_1}}\big)$ and

$$\bar{\Lambda}_2 = \sum_{i=1}^{N}\frac{(i+t_2)^{s_2} H_{N,s_2,t_2}}{(i+t_1)^{s_1} H_{N,s_1,t_1}} f\left(\frac{(i+t_2)^{s_2} H_{N,s_2,t_2}}{(i+t_1)^{s_1} H_{N,s_1,t_1}}\right)$$
$$- \sum_{i=1}^{N}\frac{(i+t_2)^{s_2} H_{N,s_2,t_2}}{(i+t_1)^{s_1} H_{N,s_1,t_1}} f\left(\frac{1}{N}\sum_{i=1}^{N}\frac{(i+t_2)^{s_2} H_{N,s_2,t_2}}{(i+t_1)^{s_1} H_{N,s_1,t_1}}\right).$$

If $t \mapsto tf(t)$ is a concave function, then the inequality signs are reversed.

Proof Since the corollary is a special case of the previous one, its proof is provided by inserting equation (3), which defines the Zipf–Mandelbrot law instead of p_i, $i = 1, \ldots, N$, as was already done for q_i. That is, $p_i = \frac{1}{(i+t_1)^{s_1} H_{N,s_1,t_1}}$, $i = 1, \ldots, N$, where $P_N = 1$. The rest of the proof follows along the same lines as in Corollary 7, so inequalities (37) and (38) follow for convex functions f and $t \mapsto tf(t)$, respectively. They change their signs in the case of concavity as a consequence of the Jensen inequality implicitly included. □

Finally, if both **p** and **q** are defined via the Zipf law (5), then the following statements hold.

Corollary 9 *Let $I \subseteq \mathbb{R}$ be an interval and suppose $N \in \mathbb{N}$, $s_1, s_2 > 0$ are such that $i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}} \in I$, $i = 1, \ldots, N$.*

If $f : I \to \mathbb{R}$ is a convex function, then

$$\frac{1}{N^{s_2} H_{N,s_2}} \Lambda_3 + f(1) \leq \tilde{D}_f(i, N, s_1, s_2) \leq \frac{1}{H_{N,s_2}} \Lambda_3 + f(1), \tag{39}$$

where

$$\Lambda_3 = \sum_{i=1}^{N} f\left(i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}}\right) - Nf\left(\frac{1}{N} \sum_{i=1}^{N} i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}}\right).$$

If f is a concave function, then the inequality signs are reversed.
If $t \mapsto tf(t)$ is a convex function, then

$$\frac{1}{N^{s_2} H_{N,s_2}} \bar{\Lambda}_3 + f(1) \leq \tilde{D}_{\mathrm{id} \cdot f}(i, N, s_1, s_2) \leq \frac{1}{H_{N,s_2}} \bar{\Lambda}_3 + f(1), \tag{40}$$

where $\tilde{D}_{\mathrm{id} \cdot f}(i, N, s_1, s_2) := \sum_{i=1}^{N} \frac{1}{i^{s_2} H_{N,s_2}} f(i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}})$ and

$$\bar{\Lambda}_3 = \sum_{i=1}^{N} i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}} f\left(i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}}\right) - \sum_{i=1}^{N} i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}} f\left(\frac{1}{N} \sum_{i=1}^{N} i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}}\right).$$

If $t \mapsto tf(t)$ is a concave function, then the inequality signs are reversed.

Proof Inequalities (39) and (40) are proved analogously to Corollary 8 if we observe the probability mass functions p_i and q_i as Zipf laws defined by (5). □

Let us provide the accompanied results of this type for some special cases of f-divergences, starting with the Kullback–Leibler divergence (10). Again, we firstly observe the more general case in which only one of two N-tuples **p** and **q** is defined via the Zipf–Mandelbrot law (3).

Corollary 10 *Let $\mathbf{p} = (p_1, \ldots, p_N)$ be an N-tuple of nonnegative real numbers with $P_N = \sum_{i=1}^{N} p_i$, $N \in \mathbb{N}$ and $s_2 > 0$, $t_2 \geq 0$.*

If the logarithm base is greater than 1, then

$$\frac{1}{(N+t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_{\mathrm{KL}}^1 \leq \mathrm{KL}(i, N, s_2, t_2, \mathbf{p}) \leq \frac{1}{(1+t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_{\mathrm{KL}}^1, \tag{41}$$

where

$$\Lambda_{\text{KL}}^1 = \sum_{i=1}^{N} p_i(i + t_2)^{s_2} H_{N,s_2,t_2} \log\big(p_i(i + t_2)^{s_2} H_{N,s_2,t_2}\big)$$

$$- \sum_{i=1}^{N} p_i(i + t_2)^{s_2} H_{N,s_2,t_2} \log\left(\frac{1}{N} \sum_{1=1}^{N} p_i(i + t_2)^{s_2} H_{N,s_2,t_2}\right).$$

If the logarithm base is less than 1, *then the inequality signs are reversed.*

Proof It follows from Corollary 7 as a special case of inequalities (36), for the function $t \mapsto t \log t$, which is convex when the logarithm base is greater than 1. It can also be derived from Corollary 1 and Remark 2 in the context of the Zipf–Mandelbrot law. □

When both **p** and **q** are defined via the Zipf–Mandelbrot law (3) or via the Zipf law (5), the following statements hold.

Corollary 11 *Let $N \in \mathbb{N}$ and $s_1, s_2 > 0$, $t_1, t_2 \geq 0$.*
If the logarithm base is greater than 1, *then*

$$\frac{1}{(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_{\text{KL}}^2 \leq \text{KL}(i, N, s_1, s_2, t_1, t_2) \leq \frac{1}{(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_{\text{KL}}^2, \tag{42}$$

where

$$\Lambda_{\text{KL}}^2 = \sum_{i=1}^{N} \frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}} \log\left(\frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}}\right)$$

$$- \sum_{i=1}^{N} \frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}} \log\left(\frac{1}{N} \sum_{1=1}^{N} \frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}}\right).$$

If the parameters $t_1, t_2 = 0$, the corresponding inequalities for the Zipf law follow:

$$\frac{1}{N^{s_2} H_{N,s_2}} \tilde{\Lambda}_{\text{KL}}^2 \leq \text{KL}(i, N, s_1, s_2) \leq \frac{1}{H_{N,s_2}} \tilde{\Lambda}_{\text{KL}}^2, \tag{43}$$

where

$$\tilde{\Lambda}_{\text{KL}}^2 = \sum_{i=1}^{N} \frac{i^{s_2-s_1} H_{N,s_2}}{H_{N,s_1}} \log\left(\frac{i^{s_2-s_1} H_{N,s_2}}{H_{N,s_1}}\right)$$

$$- \sum_{i=1}^{N} \frac{i^{s_2-s_1} H_{N,s_2}}{H_{N,s_1}} \log\left(\frac{1}{N} \sum_{1=1}^{N} \frac{i^{s_2-s_1} H_{N,s_2}}{H_{N,s_1}}\right).$$

If the logarithm base is less than 1, *then the signs in inequalities* (42) *and* (43) *are reversed.*

Proof Inequalities (42) follow from Corollary 8 as a special case of inequalities (38), for the function $t \mapsto t \log t$, which is convex when the logarithm base is greater than 1.

Similarly, inequalities (43) follow from Corollary 9 as a special case of inequalities (40). □

The following corollaries deal with the Hellinger distance (11) considering one or two N-tuples defined via the Zipf–Mandelbrot law or the Zipf law, as its special case.

Corollary 12 *Let* $\mathbf{p} = (p_1, \ldots, p_N)$ *be an N-tuple of nonnegative real numbers with* $P_N = \sum_{i=1}^{N} p_i$, $N \in \mathbb{N}$ *and* $s_2 > 0$, $t_2 \geq 0$.
Then

$$\frac{1}{2(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_h^1 \leq h^2(i, N, s_2, t_2, \mathbf{p}) \leq \frac{1}{2(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_h^1, \tag{44}$$

where

$$\Lambda_h^1 = \sum_{i=1}^{N} \left(1 - \sqrt{p_i(i + t_2)^{s_2} H_{N,s_2,t_2}}\right)^2 - N\left(1 - \sqrt{\frac{1}{N} \sum_{i=1}^{N} p_i(i + t_2)^{s_2} H_{N,s_2,t_2}}\right)^2.$$

Proof It follows from Corollary 7 as a special case of inequalities (35), for the convex function $t \mapsto \frac{1}{2}(\sqrt{t} - 1)^2$. It can also be derived from Corollary 2 and Remark 4 in the context of the Zipf–Mandelbrot law. □

When both $\mathbf{p}$ and $\mathbf{q}$ are defined via the Zipf–Mandelbrot law (3) or via the Zipf law (5), the following statements hold.

Corollary 13 *Let* $N \in \mathbb{N}$ *and* $s_1, s_2 > 0$, $t_1, t_2 \geq 0$.
Then

$$\frac{1}{2(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_h^2 \leq h^2(i, N, s_1, s_2, t_1, t_2) \leq \frac{1}{2(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_h^2, \tag{45}$$

where

$$\Lambda_h^2 = \sum_{i=1}^{N} \left(1 - \sqrt{\frac{(i + t_1)^{s_1} H_{N,s_1,t_1}}{(i + t_2)^{s_2} H_{N,s_2,t_2}}}\right)^2 - N\left(1 - \sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{(i + t_1)^{s_1} H_{N,s_1,t_1}}{(i + t_2)^{s_2} H_{N,s_2,t_2}}}\right)^2.$$

If parameters $t_1, t_2 = 0$, *the corresponding inequalities for the Zipf law follow*:

$$\frac{1}{2N^{s_2} H_{N,s_2}} \tilde{\Lambda}_h^2 \leq h^2(i, N, s_1, s_2) \leq \frac{1}{2H_{N,s_2}} \tilde{\Lambda}_h^2, \tag{46}$$

where

$$\tilde{\Lambda}_h^2 = \sum_{i=1}^{N} \left(1 - \sqrt{i^{s_1 - s_2} \frac{H_{N,s_1}}{H_{N,s_2}}}\right)^2 - N\left(1 - \sqrt{\frac{1}{N} \sum_{i=1}^{N} i^{s_1 - s_2} \frac{H_{N,s_1}}{H_{N,s_2}}}\right)^2.$$

Proof Inequalities (45) follow from Corollary 8 as a special case of inequalities (37), for the convex function $t \mapsto \frac{1}{2}(\sqrt{t} - 1)^2$.

Similarly, inequalities (46) follow from Corollary 9 as a special case of inequalities (39). □

In the sequel we provide the results of this type for the Bhattacharyya coefficient (13), starting with one N-tuple defined via the Zipf–Mandelbrot law and proceeding with both such N-tuples, as well as with the Zipf law, as its special case.

Corollary 14 *Let* $\mathbf{p} = (p_1, \ldots, p_N)$ *be an N-tuple of nonnegative real numbers with* $P_N = \sum_{i=1}^{N} p_i$, $N \in \mathbb{N}$ *and* $s_2 > 0$, $t_2 \geq 0$.
Then

$$1 - \frac{1}{(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_B^1 \geq B(i, N, s_2, t_2, \mathbf{p},) \geq 1 - \frac{1}{(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_B^1, \qquad (47)$$

where

$$\Lambda_B^1 = \sqrt{N \sum_{i=1}^{N} p_i(i + t_2)^{s_2} H_{N,s_2,t_2}} - \sum_{i=1}^{N} \sqrt{p_i(i + t_2)^{s_2} H_{N,s_2,t_2}}.$$

Proof It follows from Corollary 7 as a special case of inequalities (35), for the convex function $t \mapsto -\sqrt{t}$. It can also be derived from Corollary 3 and Remark 5 in the context of the Zipf–Mandelbrot law. □

Corollary 15 *Let* $N \in \mathbb{N}$ *and* $s_1, s_2 > 0$, $t_1, t_2 \geq 0$.
Then

$$1 - \frac{1}{(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_B^2 \geq B(i, N, s_1, t_1, s_2, t_2) \geq 1 - \frac{1}{(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_B^2, \qquad (48)$$

where

$$\Lambda_B^2 = \sqrt{N \sum_{i=1}^{N} \frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}}} - \sum_{i=1}^{N} \sqrt{\frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}}}.$$

If parameters $t_1, t_2 = 0$, *the corresponding inequalities for the Zipf law follow*:

$$1 - \frac{1}{N^{s_2} H_{N,s_2}} \tilde{\Lambda}_B^2 \geq B(i, N, s_1, s_2) \geq 1 - \frac{1}{H_{N,s_2}} \tilde{\Lambda}_B^2, \qquad (49)$$

where

$$\tilde{\Lambda}_B^2 = \sqrt{N \sum_{i=1}^{N} i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}}} - \sum_{i=1}^{N} \sqrt{i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}}}.$$

Proof Inequalities (48) follow from Corollary 8 as a special case of inequalities (37), for the convex function $t \mapsto -\sqrt{t}$.

Similarly, inequalities (49) follow from Corollary 9 as a special case of inequalities (39). □

In the same manner we proceed with analogous results for the chi-square divergence (14) and the total variation distance (15).

Corollary 16 *Let* $\mathbf{p} = (p_1, \ldots, p_N)$ *be an N-tuple of real numbers with* $P_N = \sum_{i=1}^{N} p_i$, $N \in \mathbb{N}$ *and* $s_2 > 0$, $t_2 \geq 0$.
Then

$$\frac{1}{(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_{\text{chi}}^1 \leq \chi^2(i, N, s_2, t_2, \mathbf{p}) \leq \frac{1}{(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_{\text{chi}}^1, \qquad (50)$$

where

$$
\Lambda_{\text{chi}}^1 = \sum_{i=1}^{N} \left(p_i (i + t_2)^{s_2} H_{N,s_2,t_2} - 1 \right)^2 - N \left(\frac{1}{N} \sum_{i=1}^{N} p_i (i + t_2)^{s_2} H_{N,s_2,t_2} - 1 \right)^2.
$$

Proof It follows from Corollary 7 as a special case of inequalities (35), for the convex function $t \mapsto (t-1)^2$. It can also be derived from Corollary 4 and Remark 6 in the context of the Zipf–Mandelbrot law. □

Corollary 17 *Let* $N \in \mathbb{N}$ *and* $s_1, s_2 > 0$, $t_1, t_2 \geq 0$.
Then

$$
\frac{1}{(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_{\text{chi}}^2 \leq \chi^2(i, N, s_1, t_1, s_2, t_2) \leq \frac{1}{(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_{\text{chi}}^2, \tag{51}
$$

where

$$
\Lambda_{\text{chi}}^2 = \sum_{i=1}^{N} \left(\frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}} - 1 \right)^2 - N \left(\frac{1}{N} \sum_{i=1}^{N} \frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}} - 1 \right)^2.
$$

If the parameters $t_1, t_2 = 0$, *the corresponding inequalities for the Zipf law follow*:

$$
\frac{1}{N^{s_2} H_{N,s_2}} \tilde{\Lambda}_{\text{chi}}^2 \leq \chi^2(i, N, s_1, s_2) \leq \frac{1}{H_{N,s_2}} \tilde{\Lambda}_{\text{chi}}^2, \tag{52}
$$

where

$$
\tilde{\Lambda}_{\text{chi}}^2 = \sum_{i=1}^{N} \left(i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}} - 1 \right)^2 - N \left(\frac{1}{N} \sum_{i=1}^{N} i^{s_2-s_1} \frac{H_{N,s_2}}{H_{N,s_1}} - 1 \right)^2.
$$

Proof Inequalities (51) follow from Corollary 8 as a special case of inequalities (37), for the convex function $t \mapsto (t-1)^2$.

Similarly, inequalities (52) follow from Corollary 9 as a special case of inequalities (39). □

Corollary 18 *Let* $\mathbf{p} = (p_1, \ldots, p_N)$ *be an N-tuple of real numbers with* $P_N = \sum_{i=1}^{N} p_i$, $N \in \mathbb{N}$ *and* $s_2 > 0$, $t_2 \geq 0$.
Then

$$
\frac{1}{(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_V^1 \leq V(i, N, s_2, t_2, \mathbf{p}) \leq \frac{1}{(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_V^1, \tag{53}
$$

where

$$
\Lambda_V^1 = \sum_{i=1}^{N} \left| p_i (i + t_2)^{s_2} H_{N,s_2,t_2} - 1 \right| - N \left| \frac{1}{N} \sum_{i=1}^{N} p_i (i + t_2)^{s_2} H_{N,s_2,t_2} - 1 \right|.
$$

Proof It follows from Corollary 7 as a special case of inequalities (35), for the convex function $t \mapsto |t-1|$. It can also be derived from Corollary 5 and Remark 7 in the context of the Zipf–Mandelbrot law. □

Corollary 19 *Let $N \in \mathbb{N}$ and $s_1, s_2 > 0$, $t_1, t_2 \geq 0$.*
Then

$$\frac{1}{(N + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_V^2 \leq V(i, N, s_1, t_1, s_2, t_2) \leq \frac{1}{(1 + t_2)^{s_2} H_{N,s_2,t_2}} \Lambda_V^2, \tag{54}$$

where

$$\Lambda_V^2 = \sum_{i=1}^{N} \left| \frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}} - 1 \right| - N \left| \frac{1}{N} \sum_{i=1}^{N} \frac{(i + t_2)^{s_2} H_{N,s_2,t_2}}{(i + t_1)^{s_1} H_{N,s_1,t_1}} - 1 \right|.$$

If parameters $t_1, t_2 = 0$, the corresponding inequalities for the Zipf law follow:

$$\frac{1}{N^{s_2} H_{N,s_2}} \tilde{\Lambda}_V^2 \leq V(i, N, s_1, s_2) \leq \frac{1}{H_{N,s_2}} \tilde{\Lambda}_V^2, \tag{55}$$

where

$$\tilde{\Lambda}_V^2 = \sum_{i=1}^{N} \left| i^{s_2 - s_1} \frac{H_{N,s_2}}{H_{N,s_1}} - 1 \right| - N \left| \frac{1}{N} \sum_{i=1}^{N} i^{s_2 - s_1} \frac{H_{N,s_2}}{H_{N,s_1}} - 1 \right|.$$

Proof Inequalities (54) follow from Corollary 8 as a special case of inequalities (37), for the convex function $t \mapsto |t - 1|$.

Similarly, inequalities (55) follow from Corollary 9 as a special case of inequalities (39). □

In order to conclude this section providing the Jensen-inequality related results for the f-divergences based on the Zipf–Mandelbrot law (or the Zipf law), for the triangular discrimination (16) we give only the latter one: the bounds obtained in the case of both N-tuples observed via the Zipf law.

Corollary 20 *Let $N \in \mathbb{N}$ and $s_1, s_2 > 0$.*
Then

$$\frac{1}{N^{s_2} H_{N,s_2}} \tilde{\Lambda}_\Delta^2 \leq \Delta(i, N, s_1, s_2) \leq \frac{1}{H_{N,s_2}} \tilde{\Lambda}_\Delta^2, \tag{56}$$

where

$$\tilde{\Lambda}_\Delta^2 = \sum_{i=1}^{N} \frac{(i^{s_2} H_{N,s_2} - i^{s_1} H_{N,s_1})^2}{i^{s_1} H_{N,s_1} (i^{s_1} H_{N,s_1} + i^{s_2} H_{N,s_2})} - N \frac{\left(\frac{1}{N} \sum_{i=1}^{N} i^{s_2 - s_1} \frac{H_{N,s_2}}{H_{N,s_1}} - 1 \right)^2}{\frac{1}{N} \sum_{i=1}^{N} i^{s_2 - s_1} \frac{H_{N,s_2}}{H_{N,s_1}} + 1}.$$

Proof The inequalities can easily be deduced from Corollary 9 as a special case of inequalities (39), for the convex function $t \mapsto \frac{(t-1)^2}{t+1}$. It can also be derived from Corollary 6 and Remark 8 in the context of the Zipf law. □

5 An application of the Zipf law

In the final section we are going to show how the experimental character of the Zipf law can be interpreted through the bounds (43) obtained for the Kullback–Leibler divergence.

Namely, the coefficients s_1 and s_2 from the Zipf law were analyzed by Gelbukh and Sidorov in [25] as assigned to the Russian and English languages. They calculated the mentioned coefficients and their difference for each of the 39 literature texts in both languages, with more than 10,000 running words inside of each of them. In the process they obtained the average of $s_1 = 0.892869$ for the Russian and $s_2 = 0.973863$ for the English language.

In this context, with the described experimental values of s_1 and s_2 involved, the bounds for the Kullback–Leibler divergence in (43) assume the following form which thus depends only on the parameter N.

Example 1 Let $\mathbf{p} = (p_1, \dots, p_N)$ and $\mathbf{q} = (q_1, \dots, q_N)$ be distributions associated to the Russian and English languages, respectively, and let $N \in \mathbb{N}$ be a parameter. If the logarithm base is greater than 1, then

$$\frac{1}{N^{0.973863} H_{N,0.973863}} \tilde{\Lambda}_{\mathrm{KL}}^2 \le \mathrm{KL}(i, N) \le \frac{1}{H_{N,0.973863}} \tilde{\Lambda}_{\mathrm{KL}}^2, \tag{57}$$

where

$$\tilde{\Lambda}_{\mathrm{KL}}^2 = \sum_{i=1}^{N} \frac{i^{0.080994} H_{N,0.973863}}{H_{N,0.892869}} \log\left(\frac{i^{0.080994} H_{N,0.973863}}{H_{N,0.892869}} \right)$$
$$- \sum_{i=1}^{N} \frac{i^{0.080994} H_{N,0.973863}}{H_{N,0.892869}} \log\left(\frac{1}{N} \sum_{1=1}^{N} \frac{i^{0.080994} H_{N,0.973863}}{H_{N,0.892869}} \right).$$

Proof It follows directly from (43) when inserting the experimental values of s_1 and s_2. □

6 Conclusions

In this paper we investigated f-divergences that originate from the Csiszár functional and their link to the Jensen inequality with a specific type of the Jensen-type interpolating inequalities. By means of these inequalities we derived new bounds for f-divergences in general via the Csiszár functional and in particular for the Kullback–Leibler divergence, Hellinger distance, Bhattacharyya distance (coefficient), χ^2-divergence, total variation distance and triangular discrimination. Consequently, we deduced analogous results in the light of the well-known Zipf–Mandelbrot law, with the adequate probability mass functions and the adjusted form of the Csiszár functional. The Zipf–Mandelbrot law was analyzed as a more general form of the Zipf law, for which we also gave the corresponding results and an application in linguistics in order to accentuate its experimental character. Thus this paper includes three important and widely investigated issues: the Jensen inequality, the divergences (for probability distributions) and the Zipf–Mandelbrot law with its less general, but not less important form, the Zipf law. In this way, the paper can be of an interest for mathematicians who investigate any of these fields with an accent put on mathematical inequalities, as well as for the interdisciplinary fields (*e.g.* linguistics was involved in this case).

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
[1]Faculty of Civil Engineering, Architecture and Geodesy, University of Split, Split, Croatia. [2]Catholic University of Croatia, Zagreb, Croatia. [3]Faculty of Textile Technology, University of Zagreb, Zagreb, Croatia. [4]RUDN University, Moscow, Russia.

Publisher's Note

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

References

1. Csiszár, I.: Information-type measures of difference of probability functions and indirect observations. Studia Sci. Math. Hung. **2**, 299–318 (1967)
2. Csiszár, I.: Information measures: 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 (1978)
3. Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)
4. Kullback, S., Leibler, R.A.: On information and sufficiency. Ann. Math. Stat. **22**(1), 79–86 (1951)
5. Dragomir, S.S.: Some inequalities for the Csiszár Φ-divergence, pp. 1–13. RGMIA (2001)
6. Taneja, I.J.: Bounds on triangular discrimination, harmonic mean and symmetric Chi-square divergences. arXiv:math/0505238
7. Mitrinović, D.S., Pečarić, J., Fink, A.M.: Classical and New Inequalities in Analysis. Kluwer Academic, Dordrecht (1993)
8. Krnić, M., Lovričević, N., Pečarić, J.: On McShane's functional's properties and its applications. Period. Math. Hung. **66**(2), 159–180 (2013)
9. Krnić, M., Lovričević, N., Pečarić, J.: Superadditivity of the Levinson functional and applications. Period. Math. Hung. **71**(2), 166–178 (2015)
10. Krnić, M., Lovričević, N., Pečarić, J.: Jessen's functional, its properties and applications. An. Ştiinţ. Univ. 'Ovidius' Constanţa, Ser. Mat. **20**(1), 225–248 (2012)
11. Dragomir, S.S., Pečarić, J., Persson, L.E.: Properties of some functionals related to Jensen's inequality. Acta Math. Hung. **70**(1–2), 129–143 (1996)
12. Krnić, M., Lovričević, N., Pečarić, J., Perić, J.: Superadditivity and Monotonicity of the Jensen-Type Functionals. Element (2015)
13. Manin, D.Y.: Mandelbrot's model for Zipf's law: can Mandelbrot's model explain Zipf's law for language? J. Quant. Linguist. **16**(3), 274–285 (2009)
14. Mandelbrot, B.: An informational theory of the statistical structure of language. In: Jackson, W. (ed.) Communication Theory, pp. 486–502. Academic Press, New York (1953)
15. Mandelbrot, B.: Information Theory and Psycholinguistics. Scientific Psychology: Principles and Approaches. Basic Books, New York (1965)
16. Montemurro, M.A.: Beyond the Zipf–Mandelbrot law in quantitative linguistics. arXiv:cond-mat/0104066
17. Egghe, L., Rousseau, R.: Introduction to Informetrics. Quantitative Methods in Library, Documentation and Information Science. Elsevier, New York (1990)
18. Silagadze, Z.K.: Citations and the Zipf–Mandelbrot law. Complex Syst. **11**, 487–499 (1997)
19. 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)
20. Manaris, B., Vaughan, D., Wagner, C.S., Romero, J., Davis, R.B.: Evolutionary music and the Zipf–Mandelbrot law: developing fitness functions for pleasant music. In: Proceedings of 1st European Workshop on Evolutionary Music and Art (EvoMUSART2003), Essex, pp. 522–534 (2003)
21. Horváth, L., Pečarić, Đ., Pečarić, J.: Estimations of f- and Rènyi divergences by using a cyclic refinement of the Jensen's inequality. https://doi.org/10.1007/s40840-017-0526-4
22. Renyi, A.: On measures of entropy and information. In: Proc. Fourth Berkeley Symp. Math. Statist. Prob., San Diego, vol. 1, pp. 522–534 (1992)
23. Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**(3), 379–423 (1948)
24. van Erven, T., Harremoës, P.: Rényi divergence and Kullback–Leibler divergence. J. Latex Class Files **6**(1), 1–24 (2007)
25. Gelbukh, A., Sidorov, G.: Zipf and heaps laws' coefficients depend on language. In: Proceedings of Conference on Intelligent Text Processing and Computational Linguistics. Mexico City (2001)