WILEY | Hindawi

Research Article

New Estimates for Csiszár Divergence and Zipf–Mandelbrot Entropy via Jensen–Mercer's Inequality

Muhammad Adil Khan [ID],[1] Zakir Husain [ID],[1] and Yu-Ming Chu [ID][2,3]

[1]*Department of Mathematics, University of Peshawar, Peshawar 25000, Pakistan*
[2]*Department of Mathematics, Huzhou University, Huzhou 313000, China*
[3]*Hunan Provincial Key Laboratory of Mathematical Modeling and Analysis in Engineering,*
 Changsha University of Science and Technology, Changsha 410114, China

Correspondence should be addressed to Yu-Ming Chu; chuyuming@zjhu.edu.cn

Jensen's inequality is one of the fundamental inequalities which has several applications in almost every field of science. In 2003, Mercer gave a variant of Jensen's inequality which is known as Jensen–Mercer's inequality. The purpose of this article is to propose new bounds for Csiszár and related divergences by means of Jensen–Mercer's inequality. Also, we investigate several new bounds for Zipf–Mandelbrot entropy. The idea of this article may further stimulate research on information theory with the help of Jensen–Mercer's inequality.

1. Introduction

In theory of inequalities, convex functions play an important role. The definition of convex function [1] is as follows.

Let $\phi: [\delta, \varepsilon] \subset \mathbb{R} \longrightarrow \mathbb{R}$ be a function, then ϕ is said to be convex if $\forall w, z \in [\delta, \varepsilon]$ and $0 \le t \le 1$, the inequality

$$\phi(tw + (1-t)z) \le t\phi(w) + (1-t)\phi(z), \qquad (1)$$

holds, and ϕ is said to be strictly convex if $\forall w \ne z$ and $t \in (0, 1)$, (1) holds strictly. If inequality (1) holds in reversed directions, then ϕ is said to be concave, and ϕ is said to be strictly concave if the inequality (1) holds strictly in reversed direction $\forall w \ne z$ and $t \in (0, 1)$.

There are several important inequalities which have been established with the help of convex functions. In 2003, Mercer [2] proved the following variant of Jensen's inequality, which is known as Jensen–Mercer's inequality:

$$\phi\left(\delta + \varepsilon - \frac{1}{\Theta_n} \sum_{\lambda=1}^{n} \theta_\lambda \kappa_\lambda\right) \le \phi(\delta) + \phi(\varepsilon) - \frac{1}{\Theta_n} \sum_{\lambda=1}^{n} \theta_\lambda \phi(\kappa_\lambda), \qquad (2)$$

where $\phi: [\delta, \varepsilon] \longrightarrow \mathbb{R}$ is a convex function, $[\delta, \varepsilon] \subset \mathbb{R}$ with $\kappa_\lambda \in [\delta, \varepsilon]$, and $\theta_\lambda \in \mathbb{R}^+ \, \forall \lambda \in \{1, 2, \ldots, n\}$ with $\sum_{\lambda=1}^{n} \theta_\lambda = \Theta_n$.

This inequality has been refined, extended, and generalized in different direction. Recently, Niezgoda [3] generalized this inequality by using the concept of majorization and doubly stochastic matrices. The author gave applications for separable sequences. For some other recent results related to Jensen's and Jensen–Mercer's inequality, we recommend [4–14].

In the remaining part of this paper, we present some basic terms in information theory which we deal in this paper. We start with divergence. Divergence measure is basically the measure of distance between two probability distributions. The concept of divergence measure is working efficiently to resolve different problems related to probability theory. A few divergence measures are reasonable relying upon the idea of the issue. Recently, Dragomir [15, 16], Jain [17, 18], and Taneja [19] have made many contributions in this field; they introduced different divergence measures, obtained their bounds, presented relations with other divergences, and discussed their properties.

Divergence measures have gigantic applications in assortment of fields, such as approximation of probability distributions [20, 21], biology [22], economics and political science [23, 24], analysis of contingency tables [25], signal processing [26, 27], color image segmentation [28], pattern recognition [29, 30], and magnetic resonance image analysis [31].

Some specific operational facts about divergence measure have been introduced, characterized, and applied in different fields, such as Bregman's ϕ-divergence [32], Csiszár's ϕ-divergence [33], Burbea–Rao's ϕ-divergence [34], and Renyi's like ϕ-divergence [35]. By suitable choice of the function ϕ, many divergence measures can be obtained from these generalized divergence. Because of its compact nature, Csiszár ϕ-divergence is one of the most important divergences, which is given as follows:

$$I_\phi(\theta, \gamma) = \sum_{\lambda=1}^{n} \gamma_\lambda \phi\left(\frac{\theta_\lambda}{\gamma_\lambda}\right), \tag{3}$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$ and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_n)$ are positive real n-tuples and $\phi: \mathbb{R}^+ \longrightarrow \mathbb{R}$ is a convex function. Here, convexity of the function ϕ ensures that $I_\phi(\theta, \gamma) \geq 0$.

By choosing appropriate convex function instead of function ϕ, many well-known divergences or distance functions can be acquired from (3), such as Hellinger, Renyi, Bhattacharyya, Chi-square, Kullback–Leibler, triangular discrimination, and Jefferys divergences. A concise presentation of these divergences is given as follows.

The observed data are approximated by the probability distribution in probability and statistics. This approximation leads to information loss. The primary purpose is to assess how much information is contained in the data. Approximating a distribution by $\gamma(\kappa)$ for which the real distribution is $\theta(\kappa)$ results in loss of data. Kullback–Leibler divergence is the deficiency of encoding the information regarding the distribution γ instead of genuine distribution θ. Kullback–Leibler divergence [36] can be acquired by choosing $\phi(\kappa) = \kappa \log \kappa$, $\kappa > 0$, in (3):

$$K(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} \theta_\lambda \log\left(\frac{\theta_\lambda}{\gamma_\lambda}\right). \tag{4}$$

Kullback–Leibler divergence is nonnegative and is zero if and only if $\theta_\lambda = \gamma_\lambda$. It satisfies the two properties of metric, but $K(\theta, \gamma) \neq K(\gamma, \theta)$, and does not obey the triangle inequality. Kullback–Leibler divergence is also called relative entropy. We can construct Shannon entropy [37] from Kullback–Leibler divergence, which is given as follows:

$$H(\boldsymbol{\theta}) = \sum_{\lambda=1}^{n} \theta_\lambda \log \frac{1}{\theta_\lambda}, \tag{5}$$

where θ_λ are positive real numbers with $\sum_{\lambda=1}^{n} \theta_\lambda = 1$. Shannon entropy has been used widely in physics, particularly in many quantum soluble systems, for example, see [38–41].

The extension of Kullback–Leibler divergence is Jefferys divergence [42]. It is the summation of Kullback–Leibler divergence in both directions. Jeffery's divergence can be obtained by selecting $\phi(\kappa) = (\kappa - 1)\log \kappa$, $\kappa > 0$, in (3):

$$J(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} (\theta_\lambda - \gamma_\lambda)\log\left(\frac{\theta_\lambda}{\gamma_\lambda}\right), \tag{6}$$

$J(\theta, \gamma) \geq 0$, and $J(\theta, \gamma) = J(\gamma, \theta)$, but does not obey the triangle inequality; therefore, it is not a metric. The uses of Jefferys divergence are similar to Kullback–Leibler divergence.

Bhattacharyya divergence [43] can be obtained by choosing $\phi(\kappa) = \sqrt{\kappa}$, $\kappa > 0$, in (3):

$$B(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} \sqrt{\theta_\lambda \gamma_\lambda}. \tag{7}$$

Bhattacharya divergence also satisfies three properties of a metric like Jefferys divergence, but does not obey the triangle inequality. Bhattacharya divergence has limited range. The limited range of this divergence makes it quite attractive for a distance comparison. Hussein et al. [44] used Bhattacharyya divergence for solving track-to-track association (TTTA) problem in space surveillance.

Hellinger divergence [45] is defined as

$$H(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} \left(\sqrt{\gamma_\lambda} - \sqrt{\theta_\lambda}\right)^2, \tag{8}$$

which corresponds to $\phi(\kappa) = (1 - \sqrt{\kappa})^2$, $\kappa > 0$, in (3). It satisfies all the properties of metric; therefore, Hellinger divergence is a proper metric. Hellinger divergence is used widely in data analysis, particularly when the objects being analogized are high-dimensional experiential probability distribution built from information data [46].

Total variational distance [47] can be deduced by choosing $\phi(\kappa) = |\kappa - 1|$, $\kappa > 0$, in (3):

$$V(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} |\theta_\lambda - \gamma_\lambda|. \tag{9}$$

This divergence is also a proper metric. Total variational distance is a basic quantity in probability and statistics. In information theory, variational distance is utilized to characterize solid typicality and asymptotic equipartition of sequences generated by sampling from a given distribution [48].

Now, by substituting $\phi(\kappa) = \kappa^s$, $\kappa > 0$, for $s > 1$ in (3), we can obtain Renyi divergence [49], which is given by

$$R(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} \theta_\lambda^s \gamma_\lambda^{1-s}. \tag{10}$$

Renyi divergence appears as an important tool to provide proofs of convergence of minimum description length and Bayesian estimators, both in parametric and nonparametric models [50]. Some other divergences are given below which can be obtained from (3).

(1) Chi − square divergence (see [47]): for $\phi(\kappa) = (\kappa - 1)^2$, $\kappa > 0$, the χ^2-divergence is given by

$$\chi^2(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} \frac{(\theta_\lambda - \gamma_\lambda)^2}{\gamma_\lambda}. \tag{11}$$

(2) Triangular discrimination (see [42]): the formula for triangular discrimination can be deduced by selecting $\phi(\kappa) = ((\kappa - 1)^2/\kappa + 1)$, $\kappa > 0$, which is given by

$$\Delta(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} \frac{(\theta_\lambda - \gamma_\lambda)^2}{(\theta_\lambda + \gamma_\lambda)}. \tag{12}$$

(3) Relative arithmetic – geometric divergence (see [42]): for $\phi(\kappa) = ((\kappa + 1)/2)\log((1 + \kappa)/2\kappa)$ where $\kappa > 0$, the relative arithmetic-geometric divergence is given by

$$G(\boldsymbol{\theta}, \gamma) = \sum_{\lambda=1}^{n} \frac{\theta_\lambda + \gamma_\lambda}{2} \log \frac{\theta_\lambda + \gamma_\lambda}{2\theta_\lambda}. \tag{13}$$

Zipf's law is one of the mandatory laws in information science and is often utilized in linguistics as well. In 1932, George Zipf [51] found that one can tally how often each word shows up in the text. Therefore, if p is the rank of the word and q is the frequency of occurrence of that word, then $q.p = c$, where c is a constant.

There are several utilizations of Zipf's law, such as in city populations [52], geology [53], and solar flare intensity [54]. For more details, see [55, 56].

In 1966, Benoit Mandelbrot [57] gave generalization of Zipf's law, known as Zipf–Mandelbrot law, which gave improvement on account of the low-rank words in corpus [58] and is given as follows:

$$g(\lambda) = \frac{c}{(\lambda + \tau)^u}, \tag{14}$$

where $\lambda < 1000$, $u, c > 0$, and $\tau \geq 0$. If $\tau = 0$, we obtain Zipf's law.

If $n \in \mathbb{Z}^+$, $\tau \in \mathbb{R}^+ \cup \{0\}$, $u > 0$, $\lambda \in \{1, 2, \ldots, n\}$, and $W_{n,\tau,u} = \sum_{\lambda=1}^{n} 1/(\lambda + \tau)^u$, then the probability mass function for Zipf–Mandelbrot law is given by

$$f(\lambda, n, \tau, u) = \frac{1/(\lambda + \tau)^u}{W_{n,\tau,u}}. \tag{15}$$

The formula for Zipf–Mandelbrot entropy is given as follows:

$$Z(W, \tau, u) = \frac{u}{W_{n,\tau,u}} \sum_{\lambda=1}^{n} \frac{\log(\lambda + \tau)}{(\lambda + \tau)^u} + \log W_{n,\tau,u}. \tag{16}$$

In 2017, Khan et al. [59] used some refinements of the Jensen inequality for convex functions and monotone convex functions to obtain inequalities for Zipf–Mandelbrot and Shannon entropy. They have used two parametric Zipf–Mandelbrot laws instead of different weights in the inequalities for Shannon entropy. As a result, different

parametric Zipf–Mandelbrot entropies have been obtained. In 2017, Naveed et al. [60] obtained different inequalities for these entropies by using some majorization type inequalities. In 2018, Khan et al. [55] obtained new estimations for these entropies by applying some refinements of the Jensen inequality and Taylor's formula. In 2019, Khalid et al. [61] also gave some results related to these entropies. For more recent results related to these entropies, see [61–71].

The purpose of this paper is to use Jensen–Mercer's inequality and to give several inequalities in information theory. We obtain bounds for Csiszár divergence by using Jensen–Mercer's inequality. Also, we give some bounds for different divergences by using particular convex functions. In addition, we establish bounds for Zipf–Mandelbrot entropy by applying Zipf–Mandelbrot laws instead of probability distributions in Kullback–Leibler and Jefferys divergences. Furthermore, we deduce new estimates for Zipf–Mandelbrot entropy associated to different parametric Zipf–Mandelbrot laws.

2. Bounds for Csiszár and Related Divergences

Theorem 1. *Let $\phi: [\delta, \varepsilon] \longrightarrow \mathbb{R}$ be a convex function. If θ_λ and $\gamma_\lambda (\lambda = 1, 2, \ldots, n)$ are positive real numbers with $\theta_\lambda/\gamma_\lambda \in [\delta, \epsilon]\varepsilon$ such that $\sum_{\lambda=1}^{n} \theta_\lambda = \Theta_n$ and $\sum_{\lambda=1}^{n} \gamma_\lambda = \Gamma_n$, then*

$$\phi\left(\delta + \varepsilon - \frac{\Theta_n}{\Gamma_n}\right) \leq \phi(\delta) + \phi(\varepsilon) - \frac{1}{\Gamma_n} I_\phi(\boldsymbol{\theta}, \gamma). \tag{17}$$

Proof. Replacing θ_λ by γ_λ and substituting $\kappa_\lambda = \theta_\lambda/\gamma_\lambda$ in (2), we obtain (17).

Theorem 2. *Let δ and ϵ be positive real numbers with $\delta < \varepsilon$. If θ_λ and $\gamma_\lambda (\lambda = 1, 2, \ldots, n)$ are positive real numbers with $\gamma_\lambda/\theta_\lambda \in [\delta, \varepsilon]$ such that $\sum_{\lambda=1}^{n} \theta_\lambda = \Theta_n$ and $\sum_{\lambda=1}^{n} \gamma_\lambda = \Gamma_n$, then*

$$\log\left(\delta + \varepsilon - \frac{\Gamma_n}{\Theta_n}\right)^{-1} \leq \log \frac{1}{\delta \varepsilon} - \frac{1}{\Theta_n} K(\boldsymbol{\theta}, \gamma). \tag{18}$$

Proof. Let $\phi(\kappa) = -\log \kappa$, $\kappa > 0$, so clearly $-\log \kappa$ is a convex function. Therefore, using (17) and interchanging γ_λ and θ_λ, we obtain

$$-\log\left(\delta + \varepsilon - \frac{\Gamma_n}{\Theta_n}\right) \leq -\log \delta - \log \varepsilon + \frac{1}{\Theta_n} \sum_{\lambda=1}^{n} \theta_\lambda \log \frac{\gamma_\lambda}{\theta_\lambda}. \tag{19}$$

Now, using the definition of Kullback–Leibler divergence, we obtain (18).

Corollary 1. *Let δ and ε be positive real numbers with $\delta < \epsilon$. If $\theta_\lambda (\lambda = 1, 2, \ldots, n)$ are positive real numbers such that $1/\theta_\lambda \in [\delta, \varepsilon]$ and $\sum_{\lambda=1}^{n} \theta_\lambda = 1$, then*

$$\log(\delta + \varepsilon - n)^{-1} \leq \log \frac{1}{\delta \varepsilon} + H(\boldsymbol{\theta}). \tag{20}$$

Proof. By choosing $\gamma_\lambda = 1$, $\lambda = 1, 2, \ldots, n$, in (18), we obtain (20).

Theorem 3. *Let δ and ε be positive real numbers with $\delta < \varepsilon$. If θ_λ and γ_λ ($\lambda = 1, 2, \ldots, n$) are positive real numbers such that $\theta_\lambda/\gamma_\lambda \in [\delta, \varepsilon]$ for $\lambda = 1, 2, \ldots, n$, $\sum_{\lambda=1}^{n} \theta_\lambda = \Theta_n$, and $\sum_{\lambda=1}^{n} \gamma_\lambda = \Gamma_n$, then*

$$\left(\delta + \varepsilon - \frac{\Theta_n}{\Gamma_n}\right)\log\left(\delta + \varepsilon - \frac{\Theta_n}{\Gamma_n}\right) \le \log \delta^\delta \varepsilon^\varepsilon - \frac{1}{\Gamma_n} K(\boldsymbol{\theta}, \gamma). \tag{21}$$

Proof. Let $\phi(\kappa) = \kappa \log \kappa$, $\kappa > 0$, so clearly $\kappa \log \kappa$ is a convex function. Therefore, using (17), we obtain (21).

Corollary 2. *Let δ and ε be positive real numbers with $\delta < \varepsilon$. If θ_λ ($\lambda = 1, 2, \ldots, n$) are positive real numbers such that $\theta_\lambda \in [\delta, \varepsilon]$ and $\sum_{\lambda=1}^{n} \theta_\lambda = 1$, then*

$$\left(\delta + \varepsilon - \frac{1}{n}\right)\log\left(\delta + \varepsilon - \frac{1}{n}\right) \le \log \delta^\delta \varepsilon^\varepsilon + \frac{1}{n} H(\boldsymbol{\theta}). \tag{22}$$

Proof. By choosing $\gamma_\lambda = 1$, $\lambda = 1, 2, \ldots, n$, in (21), we obtain (22).

Theorem 4. *Let all the hypotheses of Theorem 3 hold, then we have the following inequality:*

$$-\sqrt{\left(\delta + \varepsilon - \frac{\Theta_n}{\Gamma_n}\right)} \le -(\sqrt{\delta} + \sqrt{\varepsilon}) + \frac{1}{\Gamma_n} B(\boldsymbol{\theta}, \gamma). \tag{23}$$

Proof. If $\phi(\kappa) = -\sqrt{\kappa}$, $\kappa > 0$, then $\phi''(\kappa) = (1/4\kappa^{3/2}) > 0$, so $\phi(\kappa)$ is a convex function for $\kappa > 0$. Therefore, using $\phi(\kappa) = -\sqrt{\kappa}$ in Theorem 1, we obtain (23).

Theorem 5. *Let all the hypotheses of Theorem 3 hold, then we have the following inequality:*

$$\left(\delta + \varepsilon - \frac{\Theta_n + \Gamma_n}{\Gamma_n}\right)\log\left(\delta + \varepsilon - \frac{\Theta_n}{\Gamma_n}\right) \le \log \delta^{\delta-1} \varepsilon^{\varepsilon-1} - \frac{1}{\Gamma_n} J(\boldsymbol{\theta}, \gamma). \tag{24}$$

Proof. If $\phi(\kappa) = (\kappa - 1)\log \kappa$, $\kappa > 0$, then $\phi''(\kappa) = (\kappa + 1/\kappa^2) > 0$, so $\phi(\kappa)$ is a convex function for $\kappa > 0$. Therefore, using $\phi(\kappa) = (\kappa - 1)\log \kappa$ in Theorem 1, we obtain (24).

Corollary 3. *Let δ and ϵ be positive real numbers with $\delta < \varepsilon$. If θ_λ ($\lambda = 1, 2, \ldots, n$) are positive real numbers such that $\theta_\lambda \in [\delta, \varepsilon]$ and $\sum_{\lambda=1}^{n} \theta_\lambda = 1$, then*

$$\left(\delta + \varepsilon - \frac{1+n}{n}\right)\log\left(\delta + \varepsilon - \frac{1}{n}\right) \le \log \delta^{\delta-1} \varepsilon^{\varepsilon-1} + \frac{1}{n}\sum_{\lambda=1}^{n}\log\theta_\lambda + \frac{1}{n} H(\boldsymbol{\theta}). \tag{25}$$

Proof. By choosing $\gamma_\lambda = 1$, $\lambda = 1, 2, \ldots, n$, in (24), we obtain

$$\left(\delta + \varepsilon - \frac{1+n}{n}\right)\log\left(\delta + \varepsilon - \frac{1}{n}\right) \le \log \delta^{\delta-1} \varepsilon^{\varepsilon-1} + \frac{1}{n}\sum_{\lambda=1}^{n}\log\theta_\lambda - \frac{1}{n}\sum_{\lambda=1}^{n}\theta_\lambda\log\theta_\lambda. \tag{26}$$

Now, using the definition of Shannon entropy, we obtain (25).

Theorem 6. *Let all the hypotheses of Theorem 3 hold, then we have the following inequality:*

$$\left(1 - \sqrt{\delta + \varepsilon - \frac{\Theta_n}{\Gamma_n}}\right)^2 \le \delta^2 + \varepsilon^2 + 2(1 - (\sqrt{\delta} + \sqrt{\varepsilon})) - \frac{1}{\Gamma_n} H(\boldsymbol{\theta}, \gamma). \tag{27}$$

Proof. If $\phi(\kappa) = (1 - \sqrt{\kappa})^2$, $\kappa > 0$, then $\phi''(\kappa) = 1/2\kappa - (\sqrt{\kappa} - 1/2\kappa^{3/2}) \ge 0$, so $\phi(\kappa)$ is a convex function for $\kappa > 0$. Therefore, using $\phi(\kappa) = (1 - \sqrt{\kappa})^2$ in Theorem 1, we obtain (27).

Theorem 7. *Let all the hypotheses of Theorem 3 hold, then we have the following inequality:*

$$\left(\delta + \varepsilon - \frac{\Theta_n}{\Gamma_n}\right)^s \le \delta^s + \varepsilon^s - \frac{1}{\Gamma_n} R(\boldsymbol{\theta}, \gamma), \tag{28}$$

where $s > 1$.

Proof. For $s > 1$, the function $\phi(\kappa) = \kappa^s$, $\kappa > 0$, is clearly a convex function. Therefore, using $\phi(\kappa) = \kappa^s$ in Theorem 1, we obtain (28).

Theorem 8. *Let all the hypotheses of Theorem 3 hold, then we have the following inequality:*

$$\left|\delta + \varepsilon - \frac{\Theta_n + \Gamma_n}{\Gamma_n}\right| \le |\delta - 1| + |\varepsilon - 1| - \frac{1}{\Gamma_n} V(\boldsymbol{\theta}, \gamma). \tag{29}$$

Proof. If $\phi(\kappa) = |\kappa - 1|$, $\kappa \in \mathbb{R}$, and we know that absolute function is always convex on $\mathbb{R}$. Therefore, using $\phi(\kappa) = |\kappa - 1|$ in Theorem 1, we obtain (29).

Theorem 9. *Let all the hypotheses of Theorem 3 hold, then we have the following inequality:*

$$\left(\delta + \varepsilon - \frac{\Theta_n + \Gamma_n}{\Gamma_n}\right)^2 \le \delta^2 + \varepsilon^2 - 2(\delta + \varepsilon - 1) - \frac{1}{\Gamma_n}\chi^2(\boldsymbol{\theta}, \gamma). \tag{30}$$

Proof. If $\phi(\kappa) = (\kappa - 1)^2$, $\kappa > 0$, then $\phi''(\kappa) = 2 > 0$, so $\phi(z)$ is a convex function for $\kappa > 0$. Therefore, using $\phi(\kappa) = (\kappa - 1)^2$ in Theorem 1, we obtain (30).

Theorem 10. *Let all the hypotheses of Theorem 3 hold, then we have the following inequality:*

$$\frac{\left(\delta + \varepsilon - (1/\Gamma_n)(\Theta_n + \Gamma_n)\right)^2}{\left(\delta + \varepsilon - (1/\Gamma_n)(\Theta_n - \Gamma_n)\right)} \le \frac{(\delta - 1)^2}{(\delta + 1)} + \frac{(\varepsilon - 1)^2}{(\varepsilon + 1)} - \frac{1}{\Gamma_n}\Delta(\boldsymbol{\theta}, \gamma). \tag{31}$$

Proof. If $\phi(\kappa) = (\kappa - 1)^2/(\kappa + 1)$, $\kappa > 0$, then $\phi''(\kappa) = 8/(\kappa + 1)^3 \ge 0$, so $\phi(\kappa)$ is a convex function for $\kappa > 0$. Therefore, using $\phi(\kappa) = (\kappa - 1)^2/(\kappa + 1)$ in Theorem 1, we obtain (31).

Theorem 11. *Let all the hypotheses of Theorem 3 hold, then we have the following inequality:*

$$\frac{1}{2}\left(\delta + \varepsilon - \frac{1}{\Gamma_n}(\Theta_n - \Gamma_n)\right)\log\left(\frac{\delta + \varepsilon - 1/\Gamma_n(\Theta_n - \Gamma_n)}{2(\delta + \varepsilon - \Theta_n/\Gamma_n)}\right)$$

$$\le \frac{\delta + 1}{2}\log\frac{1 + \delta}{2\delta} + \frac{\varepsilon + 1}{2}\log\frac{1 + \varepsilon}{2\varepsilon} - \frac{1}{\Gamma_n}G(\boldsymbol{\theta}, \gamma). \tag{32}$$

Proof. If $\phi(\kappa) = (\kappa + 1/2)\log(1 + \kappa/2\kappa)$, $\kappa > 0$, then $\phi''(\kappa) = 1/2\kappa^2(\kappa + 1) > 0$, so $\phi(\kappa)$ is a convex function for $\kappa > 0$. Therefore, using $\phi(\kappa) = (\kappa - 1)^2/(\kappa + 1)$ in Theorem 1, we obtain (32).

3. Bounds for Zipf–Mandelbrot Entropy

In this section, we present some bounds for Zipf–Mandelbrot entropy.

Theorem 12. *Let δ and ε be positive real numbers with $\delta < \varepsilon$. If $n \in \{1, 2, 3, \ldots\}$, $\tau \ge 0$, $u > 0$, and $\gamma_\lambda > 0$, $\lambda = 1, 2, \ldots, n$, such that $\gamma_\lambda(\lambda + \tau)^u W_{n,\tau,u} \in [\delta, \varepsilon]$ and $\sum_{\lambda=1}^n \gamma_\lambda = \Gamma_n$, then*

$$\log(\delta + \varepsilon - \Gamma_n)^{-1} \le \log\frac{1}{\delta\varepsilon} + \sum_{\lambda=1}^n \frac{\log\gamma_\lambda}{(\lambda + \tau)^u W_{n,\tau,u}} + Z(W, \tau, u). \tag{33}$$

Proof. Let $\theta_\lambda = 1/(\lambda + \tau)^u W_{n,\tau,u}$, $\lambda = 1, 2, \ldots, n$, then

$$\sum_{\lambda=1}^n \theta_\lambda \log\frac{\theta_\lambda}{\gamma_\lambda} = \sum_{\lambda=1}^n \frac{1}{(\lambda + \tau)^u W_{n,\tau,u}}\log\left(\left(\frac{1}{\gamma_\lambda}\right)\left(\frac{1}{(\lambda + \tau)^u W_{n,\tau,u}}\right)\right)$$

$$= -\sum_{\lambda=1}^n \frac{\log\gamma_\lambda}{(\lambda + \tau)^u W_{n,\tau,u}} - \sum_{\lambda=1}^n \frac{\log\left((\lambda + \tau)^u W_{n,\tau,u}\right)}{(\lambda + \tau)^u W_{n,\tau,u}}$$

$$= -\sum_{\lambda=1}^n \frac{\log\gamma_\lambda}{(\lambda + \tau)^u W_{n,\tau,u}} - \sum_{\lambda=1}^n \frac{u}{(\lambda + \tau)^u W_{n,\tau,u}}\log(\lambda + \tau)$$

$$- \sum_{\lambda=1}^n \frac{\log W_{n,\tau,u}}{(\lambda + \tau)^u W_{n,\tau,u}}$$

$$= -\sum_{\lambda=1}^n \frac{\log\gamma_\lambda}{(\lambda + \tau)^u W_{n,\tau,u}} - Z(W, \tau, u). \tag{34}$$

As $W_{n,\tau,u} = \sum_{\lambda=1}^n 1/(\lambda + \tau)^u$, $\sum_{\lambda=1}^n \theta_\lambda = \sum_{\lambda=1}^n (1/(\lambda + \tau)^u W_{n,\tau,u}) = 1$. Hence, using (18) for $\theta_\lambda = 1/(\lambda + \tau)^u W_{n,\tau,u}$, $\lambda = 1, 2, \ldots, n$, we obtain (33).

The following corollary is the special case of Theorem 12.

Corollary 4. *Let δ and ε be positive real numbers with $\delta < \varepsilon$. If $n \in \{1, 2, 3, \ldots\}$, $\tau \ge 0$, $u > 0$, and $(\lambda + \tau)^u W_{n,\tau,u} \in [\delta, \varepsilon]$, then*

$$\log(\delta + \varepsilon - n)^{-1} \le \log\frac{1}{\delta\varepsilon} + Z(W, \tau, u). \tag{35}$$

Proof. By choosing $\gamma_\lambda = 1, \lambda = 1, 2, \ldots, n$, in (33), we obtain (35).

The next result gives the inequality for Zipf–Mandelbrot entropy using two different parameters.

Theorem 13. *Let δ and ε be positive real numbers with $\delta < \varepsilon$. Let $n \in \{1, 2, \ldots\}$, $\tau_1, \tau_2 \ge 0$, $u_1, u_2 > 0$, and $\left((\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}\right)/\left((\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}\right) \in [\delta, \varepsilon]$, then*

$$\log(\delta + \varepsilon - 1)^{-1} \le \log\frac{1}{\delta\varepsilon} + Z(W, \tau_1, u_1) - \sum_{\lambda=1}^n \frac{\log(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}. \tag{36}$$

Proof. Let $\theta_\lambda = 1/(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}$ and $\gamma_\lambda = 1/(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}$, $\lambda = 1, 2, \ldots, n$, then

$$\sum_{\lambda=1}^n \theta_\lambda \log\frac{\theta_\lambda}{\gamma_\lambda} = \sum_{\lambda=1}^n \frac{1}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}\log\left(\frac{(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}\right)$$

$$= -\sum_{\lambda=1}^n \frac{\log(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}} + \sum_{\lambda=1}^n \frac{\log(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}$$

$$= -Z(W, \tau_1, u_1) + \sum_{\lambda=1}^n \frac{\log(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}. \tag{37}$$

Also, $\sum_{\lambda=1}^n \theta_\lambda = \sum_{\lambda=1}^n 1/(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1} = 1$ and $\sum_{\lambda=1}^n \gamma_\lambda = \sum_{\lambda=1}^n 1/(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2} = 1$; therefore, using (18) for $\theta_\lambda = 1/(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}$ and $\gamma_\lambda = 1/(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}$, $\lambda = 1, 2, \ldots, n$, we obtain (36).

Theorem 14. *Let δ and ε be positive real numbers with $\delta < \varepsilon$. If $n \in \{1, 2, 3, \ldots\}$, $\tau \ge 0$, $u > 0$, and $\gamma_\lambda > 0$, $\lambda = 1, 2, \ldots, n$, such that $1/\gamma_\lambda(\lambda + \tau)^u W_{n,\tau,u} \in [\delta, \varepsilon]$ and $\sum_{\lambda=1}^n \gamma_\lambda = \Gamma_n$, then*

$$\left(\delta + \varepsilon - \frac{1}{\Gamma_n}\right)\log\left(\delta + \varepsilon - \frac{1}{\Gamma_n}\right) \le \log\delta^\delta\varepsilon^\varepsilon$$

$$+ \frac{1}{\Gamma_n}\sum_{\lambda=1}^n \frac{\log\gamma_\lambda}{(\lambda + \tau)^u W_{n,\tau,u}} + \frac{1}{\Gamma_n}Z(W, \tau, u). \tag{38}$$

Proof. Substituting $\theta_\lambda = 1/(\lambda + \tau)^u W_{n,\tau,u}$, $\lambda = 1, 2, \ldots, n$, in (21) and using similar method which we used in the proof of Theorem 12, we obtain (38).

The following corollary is the special case of Theorem 14.

Corollary 5. *Let δ and ε be positive real numbers with $\delta < \varepsilon$. If $n \in \{1, 2, 3, \ldots\}$, $\tau \geq 0$, $u > 0$, and $1/(\lambda + \tau)^u W_{n,\tau,u} \in [\delta, \varepsilon]$, then*

$$\left(\delta + \varepsilon - \frac{1}{n}\right)\log\left(\delta + \varepsilon - \frac{1}{n}\right) \leq \log \delta^\delta \varepsilon^\varepsilon + \frac{1}{n} Z(W, \tau, u).$$
(39)

Proof. By choosing $\gamma_\lambda = 1, \lambda = 1, 2, \ldots, n$, in (38), we obtain (39).

The next result gives the inequality for Zipf–Mandelbrot entropy using two different parameters.

Theorem 15. *Let δ and ϵ be positive real numbers with $\delta < \epsilon$. Let $n \in \{1, 2, \ldots\}, \tau_1, \tau_2 \geq 0$, $u_1, u_2 > 0$, and $\left((\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}\right)/\left((\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}\right) \in [\delta, \varepsilon]$, then*

$$(\delta + \varepsilon - 1)\log(\delta + \varepsilon - 1) \leq \log \delta^\delta \varepsilon^\varepsilon + Z(W, \tau_1, u_2)$$
$$- \sum_{\lambda=1}^{n} \frac{\log (\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}.$$
(40)

Proof. Substituting $\theta_\lambda = 1/(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}$ and $\gamma_\lambda = 1/(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}$, $\lambda = 1, 2, \ldots, n$, in (21) and using similar method which we used in the proof of Theorem 13, we obtain (40).

Theorem 16. *Let all the hypotheses of Theorem 14 hold, then we have the following inequality:*

$$\left(\delta + \varepsilon - \frac{1 + \Gamma_n}{\Gamma_n}\right)\log\left(\delta + \varepsilon - \frac{1}{\Gamma_n}\right) \leq \log \delta^{\delta-1} \varepsilon^{\varepsilon-1}$$
$$+ \frac{1}{\Gamma_n}\left[\sum_{\lambda=1}^{n} \frac{\log \gamma_\lambda}{(\lambda + \tau)^u W_{n,\tau,u}} + Z(W, \tau, u)\right.$$
$$\left. - \sum_{\lambda=1}^{n} \gamma_\lambda \log \gamma_\lambda - u \sum_{\lambda=1}^{n} \gamma_\lambda \log(\lambda + \tau)\right] - \log W_{n,\tau,u}.$$
(41)

Proof. Let $\theta_\lambda = 1/(\lambda + \tau)^u W_{n,\tau,u}$, $\lambda = 1, 2, \ldots, n$, then

$$\sum_{\lambda=1}^{n} \theta_\lambda \log \frac{\theta_\lambda}{\gamma_\lambda} = -\sum_{\lambda=1}^{n} \frac{\log \gamma_\lambda}{(\lambda + \tau)^u W_{n,\tau,u}} - Z(W, \tau, u),$$
(42)

$$\sum_{\lambda=1}^{n} \gamma_\lambda \log \frac{\theta_\lambda}{\gamma_\lambda} = -\sum_{\lambda=1}^{n} \gamma_\lambda \log \gamma_\lambda (\lambda + \tau) W_{n,\tau,u}$$
$$= -\sum_{\lambda=1}^{n} \gamma_\lambda \log \gamma_\lambda - u \sum_{\lambda=1}^{n} \gamma_\lambda \log(\lambda + \tau)$$
$$- \Gamma_n \log W_{n,\tau,u}.$$
(43)

Also, $\sum_{\lambda=1}^{n} \theta_\lambda = \sum_{\lambda=1}^{n} 1/(\lambda + \tau)^u W_{n,\tau,u} = 1$; therefore, using (42) and (43) in (24) for $\theta_\lambda = 1/(\lambda + \tau)^u W_{n,\tau,u}$, $\lambda = 1, 2, \ldots, n$, we obtain (41).

The following corollary is the particular case of Theorem 16.

Corollary 6. *Let δ and ε be positive real numbers with $\delta < \varepsilon$. If $n \in \{1, 2, 3, \ldots\}$, $\tau \geq 0$, $u > 0$, and $1/(\lambda + \tau)^u W_{n,\tau,u} \in [\delta, \varepsilon]$, then*

$$\left(\delta + \varepsilon - \frac{1 + n}{n}\right)\log\left(\delta + \varepsilon - \frac{1}{n}\right) \leq \log \delta^{\delta-1} \varepsilon^{\varepsilon-1}$$
$$+ \frac{1}{n}\left[Z(W, \tau, u) - u \sum_{\lambda=1}^{n} \log(\lambda + \tau)\right] - \log W_{n,\tau,u}.$$
(44)

Proof. By choosing $\gamma_\lambda = 1, \lambda = 1, 2, \ldots, n$, in (41), we obtain (44).

Theorem 17. *Let all the hypotheses of Theorem 15 hold, then we have the following inequality:*

$$(\delta + \varepsilon - 2)\log(\delta + \varepsilon - 1) \leq \log \delta^{\delta-1} \varepsilon^{\varepsilon-1} + Z(W, \tau_1, u_1)$$
$$- Z(W, \tau_2, u_2) + \sum_{\lambda=1}^{n}\left[\frac{\log (\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}{(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}\right.$$
$$\left. - \frac{\log (\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}\right].$$
(45)

Proof. Let $\theta_\lambda = 1/(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}$ and $\gamma_\lambda = 1/(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}$, $\lambda = 1, 2, \ldots, n$, then

$$\sum_{\lambda=1}^{n} \theta_\lambda \log \frac{\theta_\lambda}{\gamma_\lambda} = -Z(W, \tau_1, u_1) + \sum_{\lambda=1}^{n} \frac{\log (\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}},$$
(46)

$$\sum_{\lambda=1}^{n} \gamma_\lambda \log \frac{\theta_\lambda}{\gamma_\lambda} = \sum_{\lambda=1}^{n} \frac{1}{(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}\log\left(\frac{(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}\right)$$
$$= \sum_{\lambda=1}^{n} \frac{\log (\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}$$
$$- \sum_{\lambda=1}^{n} \frac{\log (\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}{(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}$$
$$= Z(W, \tau_2, u_2) - \sum_{\lambda=1}^{n} \frac{\log (\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}}{(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}}.$$
(47)

Also, $\sum_{\lambda=1}^{n} \theta_\lambda = \sum_{\lambda=1}^{n} 1/(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1} = 1$ and $\sum_{\lambda=1}^{n} \gamma_\lambda = \sum_{\lambda=1}^{n} 1/(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2} = 1$; therefore, using (46) and (47) in (24) for $\theta_\lambda = 1/(\lambda + \tau_1)^{u_1} W_{n,\tau_1,u_1}$ and $\gamma_\lambda = 1/(\lambda + \tau_2)^{u_2} W_{n,\tau_2,u_2}$, $\lambda = 1, 2, \ldots, n$, we obtain (45).

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

M. Adil Khan carried out the proof of Theorems 1– 5 and Corollaries 1– 3 and drafted the manuscript. Husain carried out the proof of Theorems 6– 12 and Corollary 4. Chu provided the main idea, carried out the proof of Theorems 13– 17 and Corollaries 5 and 6, completed the final revision, and submitted the article. All authors read and approved the final manuscript.

Acknowledgments

References

[1] J. Pečarić, F. Proschan, and Y. L. Tong, *Convex Functions, Partial Orderings, and Statistical Applications*, Academic Press, Boston, MA, USA, 1992.

[2] A. McD. Mercer, "A variant of Jensens inequality," *Journal of Inequalities in Pure and Applied*, vol. 73, no. 4, pp. 1–4, 2003.

[3] M. Niezgoda, "A generalization of Mercer's result on convex functions," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 71, no. 7-8, pp. 2771–2779, 2009.

[4] S. Khan, M. Adil Khan, and Y. M. Chu, "Converses of the Jensen inequality derived from the Green functions with applications in information theory," *Mathematical Methods in the Applied Sciences*, vol. 43, no. 5, pp. 2577–2587, 2020.

[5] M. Adil Khan, Đ. Pečarić, and J. Pečarić, "New refinement of the Jensen inequality associated to certain functions with applications," *Journal of Inequalities and Applications*, vol. 2020, Article ID 76, 11 pages, 2020.

[6] M. Adil Khan, A. R. Khan, and J. Pečarić, "On the refinements of Jensen-Mercer's inequality," *Journal of Numerical Analysis and Approximation Theory*, vol. 41, no. 1, pp. 62–81, 2012.

[7] K. A. Khan, M. Adil Khan, and U. Sadaf, "New refinement of Jensen-Mercer's operator inequality and applications to means," *Punjab University Journal of Mathematics*, vol. 49, no. 2, pp. 127–151, 2017.

[8] M. Adil Khan, J. Khan, and J. Pečarić, "Generalization of Jensen's and Jensen-Steffensen's inequalities by generalized majorization theorem," *Journal of Mathematical Inequalities*, vol. 11, no. 4, pp. 1049–1074, 2017.

[9] H. R. Moradi, M. E. Omidvar, M. Adil Khan, and K. Nikodem, "Around Jensen's inequality for strongly convex functions," *Aequationes Mathematicae*, vol. 92, no. 1, pp. 25–37, 2018.

[10] S. Zaheer Ullah, M. Adil Khan, and Y.-M. Chu, "A note on generalized convex functions," *Journal of Inequalities and Applications*, vol. 2019, Article ID 291, 10 pages, 2019.

[11] M. Adil Khan, J. Pečarić, J. Pečarić, and Y.-M. Chu, "Refinements of Jensen's and McShane's inequalities with applications," *AIMS Mathematics*, vol. 5, no. 5, pp. 4931–4945, 2020.

[12] S. Z. Ullah, M. A. Khan, Z. A. Khan, and Y. Chu, "Coordinate strongly s-convex functions and related results," *Journal of Mathematical Inequalities*, vol. 14, no. 3, pp. 829–843, 2020.

[13] M. A. Khan, Đ. Pečarić, and J. Pečarić, "A new refinement of the Jensen inequality with applications in information theory," *Bulletin of the Malaysian Mathematical Sciences Society*, pp. 1–12, 2020.

[14] M. Adil Khan, S. Khan, I. Ullah, K. Ali Khan, and Y. M. Chu, "A novel approach to the Jensen gap through Taylor's theorem," *Mathematical Methods in the Applied Sciences*, pp. 1–10, 2020.

[15] S. S. Dragomir, J. Sunde, and C. Buse, "New inequalities for Jeffreys divergence measure," *Tamsui Oxford Journal of Information and Mathematical Sciences*, vol. 16, no. 2, pp. 295–309, 2000.

[16] S. S. Dragomir, "Some inequalities for the csiszár Φ-divergence," *Italian Journal of Pure and Applied Mathematics*, vol. 2, no. 17, pp. 15–28, 2004.

[17] K. C. Jain and R. Mathur, "A symmetric divergence measure and its bounds," *Tamkang Journal of Mathematics*, vol. 42, no. 4, pp. 493–503, 2011.

[18] K. C. Jain and P. Chhabra, "Bounds on nonsymmetric divergence measure in terms of other symmetric and nonsymmetric divergence measures," *International Scholarly Research Notices*, vol. 2014, Article ID 820375, 9 pages, 2014.

[19] I. J. Taneja, "Bounds on non symmetric divergence measures in terms of symmetric divergence measures," *Journal of Combinatorics, Information & System Sciences*, vol. 29, no. 14, pp. 115–134, 2005.

[20] D. Kazakos and T. Cotsidas, "A decision theory approach to the approximation of discrete probability densities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 1, pp. 61–67, 1980.

[21] C. Chow and C. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467, 1968.

[22] E. C. Pielou, *Ecological Diversity*, Wiley, New York, NY, USA, 1975.

[23] H. Theil, *Statistical Decomposition Analysis*, North-Holland, Amsterdam, Netherlands, 1972.

[24] H. Theil, *Economics and Information Theory*, North-Holland, Amsterdam, Netherlands, 1967.

[25] D. V. Gokhale and S. Kullback, *Information in Contingency Tables*, Marcel Dekker, New York, NY, USA, 1978.

[26] T. T. Kadota and L. A. Shepp, "On the best finite set of linear observables for discriminating two Gaussian signals," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 288–294, 1967.

[27] T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Transactions on Communications*, vol. 20, no. 3, pp. 52–60, 1967.

[28] F. Nielsen and S. Boltz, "The burbea-rao and Bhattacharyya centroids," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 5455–5466, 2011.

[29] L. K. Jones and C. L. Byrne, "General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 1, pp. 23–30, 1990.

[30] H. C. Chen, *Statistical Pattern Recognition*, Hoyderc Book Co., Rocelle Park, NJ, USA, 1973.

[31] B. Vemuri, M. Liu, S. Amari, and F. Nielsen, "Total Bregman divergence and its applications to DTI analysis," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 475–483, 2010.

[32] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 3, pp. 200–217, 1967.

[33] I. Csiszár, "Information type measures of differences of probability distribution and indirect observations," *Studia*

Scientiarum Mathematicarum Hungarica, vol. 2, pp. 299–318, 1967.

[34] J. Burbea and C. Rao, "On the convexity of some divergence measures based on entropy functions," *IEEE Transactions on Information Theory*, vol. 28, no. 3, pp. 489–495, 1982.

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

[36] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[37] C. E. Shannon, "A mathematical theory of communications," *Bell Labs Technical Journal*, vol. 27, pp. 623–659, 1958.

[38] G.-H. Suna, M. Avila Aokia, and S.-H. Dongb, "Quantum information entropies of the eigenstates for the PöschlTeller-like potential," *Chinese Physics B*, vol. 22, no. 5, Article ID 050302, 2013.

[39] G. Yánẽz-Navarro, G.-H. Sun, T. Dytrich, K. D. Launey, S.-H. Dong, and J. P. Jerry, "Quantum information entropies for position-dependent mass Schrödinger problem," *Annals of Physics*, vol. 348, pp. 153–160, 2014.

[40] G.-H. Sun, S.-H. Dong, K. D. Launey, T. Dytrych, and J. P. Draayer, "Shannon information entropy for a hyperbolic double-well potential," *International Journal of Quantum Chemistry*, vol. 115, no. 14, pp. 891–899, 2015.

[41] R. Valencia-Torres, G.-H. Sun, and S.-H. Dong, "Quantum information entropy for a hyperbolical potential function," *Physica Scripta*, vol. 90, no. 3, Article ID 035205, 2015.

[42] K. C. Jain and P. Chhabra, "New series of information divergence measures and their properties," *Applied Mathematics & Information Sciences*, vol. 10, no. 4, pp. 1433–1446, 2016.

[43] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distribution," *Bulletin of the Calcutta Mathematical Society*, vol. 35, p. 99110, 1943.

[44] I. I. Hussein, C. W. Roscoe, M. P. Wilkins, and P. W. Schumacher Jr, "Track-to-track association using Bhattacharyya divergence," in *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, Wailea, HI, USA, September 2015.

[45] R. Beran, "Minimum Hellinger distance estimates for parametric models," *Annals of Statistics*, vol. 5, no. 1, pp. 445–463, 1977.

[46] A. Abdullah, R. Kumar, A. McGregor, S. Vassilvitskii, Venkatasubramanian, and S. Sketching, "Embedding and dimensionality reduction in information theoretic spaces," *Artificial Intelligence Growth Statistics*, vol. 41, pp. 948–956, 2016.

[47] J. N. Kapur, "A comparative assessment of various measures of directed divergence," *Journal of Management Studies*, vol. 3, no. 1, pp. 1–16, 1984.

[48] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Hoboken, NJ, USA, 1991.

[49] A. Renyi, "On Measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, CA, USA, 1961.

[50] T. Zhang, "From ε -entropy to KL-entropy: analysis of minimum information complexity density estimation," *The Annals of Statistics*, vol. 34, no. 5, pp. 2180–2210, 2006.

[51] G. K. Zipf, "Human behavior and the principle of least effort: an introduction to human ecology," *Social Forces*, vol. 26, no. 3, pp. 340-341, 1950.

[52] K. Giesen and J. Südekum, "Zipf's law for cities in the regions and the country," *Journal of Economic Geography*, vol. 11, no. 4, pp. 667–686, 2010.

[53] D. F. Merriam, L. J. Drew, and J. H. Schuenemeyer, "Zipf?s law: a viable geological paradigm?" *Natural Resources Research*, vol. 13, no. 4, pp. 265–271, 2004.

[54] M. Newman, "Power laws, Pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.

[55] M. Adil Khan, M. Z. Al-sahwi, and Y.-M. Chu, "New estimations for Shannon and zipf-mandelbrot entropies," *Entropy*, vol. 20, no. 8, 2018.

[56] D. F. Merriam and J. C. Davis, "Using zipf's law to predict future earthquakes in Kansas," *Transactions of the Kansas Academy of Science*, vol. 112, no. 1/2, pp. 127–129, 2009.

[57] B. Mandelbrot, "Information theory and psycholinguistics: a theory of words frequencies," in *Reading in Mathematical Social Science*, P. Lazafeld and N. Henry, Eds., pp. 350–368, MIT Press, Cambridge, MA, USA, 1966.

[58] M. A. Montemurro, "Beyond the Zipf-Mandelbrot law in quantitative linguistics," *Journal of Physics A*, vol. 300, no. 3-4, pp. 567–578, 2001.

[59] M. Adil Khan, Đ. Pečaric, and J. Pečarić, "Bounds for Shannon and zipf-mandelbrot entropies," *Mathematical Methods in the Applied Sciences*, vol. 40, no. 18, pp. 7316–7322, 2017.

[60] N. Latif, Đ. Pečarić, and J. Pečarić, "Majorization, Csiszár divergence and zipf-mandelbrot law," *Journal of Inequalities and Applications*, vol. 2017, Article ID 197, 2017.

[61] S. Khalid, Đ. Pečaric, and J. Pečarić, "On Shannon and Zipf-Mandelbrot entropies and related results," *Journal of Inequalities and Applications*, vol. 2019, Article ID 99, 2019.

[62] S. Khan, M. Adil Khan, and Y.-M. Chu, "New converses of Jensen inequality via Green functions with applications," *RACSAM*, vol. 114, no. 3, Article ID 114, 2020.

[63] M. Adil Khan, S. Khan, and Y. Chu, "A new bound for the Jensen gap with applications in information theory," *IEEE Access*, vol. 8, pp. 98001–98008, 2020.

[64] S. Khan, M. Adil Khan, S. I. Butt, and Y.-M. Chu, "A new bound for the Jensen gap pertaining twice differentiable functions with applications," *Advances in Difference Equations*, vol. 2020, Article ID 333, 11 pages, 2020.

[65] K. Ahmad, M. Adil Khan, S. Khan, A. Ali, and Y.-M. Chu, "New estimates for generalized Shannon and Zipf-Mandelbrot entropies via convexity results," *Results in Physics*, vol. 18, p. 103305, 2020.

[66] M. Adil Khan, S. Khan, and Y.-M. Chu, "New estimates for the Jensen gap using s-convexity with applications," *Frontier in Physics*, vol. 8, Article ID 313, 8 pages, 2020.

[67] M. Adil Khan, M. Hanif, Z. A. Khan, K. Ahmad, and Y.-M. Chu, "Association of Jensen inequality for *s*-convex function," *Journal of Inequalities and Applications*, vol. 2019, Article ID 162, 14 pages, 2019.

[68] M. Adil Khan, Đ. Pečarić, and J. Pečarić, "Bounds for Csiszár divergence and hybrid zipf-mandelbrot entropy," *Mathematical Methods in the Applied Sciences*, vol. 42, no. 18, pp. 7411–7424, 2019.

[69] M. Adil Khan, Đ. Pečarić, and J. Pečarić, "On Zipf-Mandelbrot entropy," *Journal of Computational and Applied Mathematics*, vol. 346, pp. 192–204, 2019.

[70] S. S. Dragomir, M. Adil Khan, and A. Abathun, "Inequalities for power series with nonnegative coefficients via a reverse of Jensen inequality," *Tamkang Journal of Mathematics*, vol. 47, no. 2, pp. 221–236, 2016.

[71] M. Adil Khan, G. Ali Khan, T. Ali, and A. Kilicman, "On the refinement of Jensen's inequality," *Applied Mathematics and Computation*, vol. 262, pp. 128–135, 2015.