

Improved bounds for the Jensen gap with applications in information theory

Khurshid Ahmad^a, Muhammad Adil Khan^b, Shahid Khan^b, Yu-Ming Chu^{c,d,*}

^a Department of Statistics, Islamia College University, Peshawar 25000 Pakistan

^b Department of Mathematics, University of Peshawar, Peshawar 25000 Pakistan

^c Department of Mathematics, Huzhou University, Huzhou 313000 China

^d Institute for Advanced Study Honoring Chen Jian Gong, Hangzhou Normal University, Hangzhou 311121 China

*Corresponding author, e-mail: chuyuming2005@126.com

Received 4 Nov 2020, Accepted 10 May 2022

Available online 10 Aug 2022

ABSTRACT: In this article, we utilize strongly convex functions to improve some upper bounds of Jensen’s gap presented in [IEEE Access 8 (2020):98001–98008, Adv Differ Equ 2020 (2020):333]. This leads us towards the improvement of some existing results around various divergences, the Hermite-Hadamard and Hölder inequalities. We also demonstrate these improvements via numerical experiments. Moreover, some new results are established for the Zipf-Mandelbrot entropy.

KEYWORDS: Jensen inequality, Hermite-Hadamard inequality, Hölder inequality, strongly convex function, Green function, Csiszár divergence

MSC2020: 26A33 26A51 26D15

INTRODUCTION

A growing interest in utilizing the mathematical inequalities to various areas of science with remarkable effect has been recorded in the last couple of decades. Mathematical inequalities for convex and generalized convex functions have many useful applications in analysis, specially in optimization theory. Among them, Jensen’s inequality is most important inequality and it provides problem solving oriented tools in various fields of science, for example, economics, engineering, physics, computer science, statistics, biology, information theory, etc. This inequality generalizes the classical notion of convexity and states that [1]:

Theorem 1 Let $\psi : [\rho_1, \rho_2] \rightarrow \mathbb{R}$ be a convex function, $\vartheta_k \in [\rho_1, \rho_2]$, $p_k \geq 0$ for $k = 1, \dots, m$ with $P_m = \sum_{k=1}^m p_k > 0$, then

$$\psi\left(\frac{1}{P_m} \sum_{k=1}^m p_k \vartheta_k\right) \leq \frac{1}{P_m} \sum_{k=1}^m p_k \psi(\vartheta_k). \quad (1)$$

The integral version of Theorem 1 in Riemann sense can be found in [2]. The inequality (1) helps to investigate the stability of time-delayed systems [3] and provides some sufficient conditions to achieve an exponential tracking performance for vehicle platoon system to improve traffic safety [4]. The Jensen’s inequality can be utilized for estimation of various divergences and the Zipf-Mandelbrot entropy [5–9]. The Jensen’s inequality has also been presented for various classes of convex functions as well for example (α, m) -convex and m -convex [10], s -convex [11], quasi-convex [12], Q -class convex [13] and strongly

convex functions [14], and some interesting generalizations of mathematical inequalities and classical convexity can be found in [15, 16].

The difference of the right and left sides of Jensen’s inequality is known as the Jensen gap. This gap gives some useful estimates, which can be useful to provide some error bounds while approximating certain parameters. This paper addresses the improvements of some upper bounds of the Jensen gap presented in [17, 18], by using strongly convex functions. This class of functions was originally introduced and studied by Polyak [19]. A strongly convex function can be defined as:

Definition 1 ([20]) For a constant $c > 0$ and an interval I , a function $\psi : I \rightarrow \mathbb{R}$ is said to be strongly convex with modulus c , if the following inequality holds

$$\psi(ts_1 + (1-t)s_2) \leq t\psi(s_1) + (1-t)\psi(s_2) - ct(1-t)(s_1 - s_2)^2, \quad (2)$$

for all $s_1, s_2 \in I$ and $t \in [0, 1]$.

From (2), the inequality $\psi(s_1) - \psi(s_2) \geq \psi'_+(s_2)(s_1 - s_2) + c(s_1 - s_2)^2$ can be verified easily. It is noteworthy that each strongly convex function is convex but the converse statement is not true generally. The idea of strong convexity actually strengthens the idea of classical convexity.

We have the following tool to check whether a function is strongly convex or not [20]:

Theorem 2 A function $\psi : I \rightarrow \mathbb{R}$ is said to be strongly convex with modulus c , if and only if a function $\Psi : I \rightarrow \mathbb{R}$, defined as $\Psi(x) = \psi(x) - cx^2$ is convex.

To construct the main results, we need the following continuous and convex Green function [21], defined on $[\rho_1, \rho_2] \times [\rho_1, \rho_2]$:

$$G_1(z, x) = \begin{cases} \rho_1 - x, & \rho_1 \leq x \leq z, \\ \rho_1 - z, & z \leq x \leq \rho_2, \end{cases} \quad (3)$$

and the following integral identity, which holds for a function $\psi \in C^2[\rho_1, \rho_2]$,

$$\psi(z) = \psi(\rho_1) + (z - \rho_1)\psi'(\rho_1) + \int_{\rho_1}^{\rho_2} G_1(z, x)\psi''(x) dx. \quad (4)$$

MAIN RESULTS

The following theorem gives a new bound for the discrete Jensen's gap. This is an improvement of the upper bound presented in [18].

Theorem 3 Let $\psi \in C^2[\rho_1, \rho_2]$ be a function such that the function $|\psi''|$ is strongly convex with modulus c . Let $\vartheta_k \in [\rho_1, \rho_2]$, $p_k \geq 0$ for $k = 1, \dots, m$ with $\sum_{k=1}^m p_k = P_m > 0$ and $\bar{\vartheta} = \frac{1}{P_m} \sum_{k=1}^m p_k \vartheta_k$, then

$$\begin{aligned} & \left| \frac{1}{P_m} \sum_{k=1}^m p_k \psi(\vartheta_k) - \psi(\bar{\vartheta}) \right| \\ & \leq \frac{\rho_2 |\psi''(\rho_1)| - \rho_1 |\psi''(\rho_2)| + c \rho_1 \rho_2 (\rho_2 - \rho_1)}{2(\rho_2 - \rho_1)} \\ & \quad \times \left(\frac{1}{P_m} \sum_{k=1}^m p_k \vartheta_k^2 - (\bar{\vartheta})^2 \right) \\ & + \frac{|\psi''(\rho_2)| - |\psi''(\rho_1)| - c(\rho_2^2 - \rho_1^2)}{6(\rho_2 - \rho_1)} \left(\frac{1}{P_m} \sum_{k=1}^m p_k \vartheta_k^3 - (\bar{\vartheta})^3 \right) \\ & \quad + \frac{c}{12} \left(\frac{1}{P_m} \sum_{k=1}^m p_k \vartheta_k^4 - (\bar{\vartheta})^4 \right). \end{aligned} \quad (5)$$

Proof: Using (4) in $\frac{1}{P_m} \sum_{k=1}^m p_k \psi(\vartheta_k)$ and $\psi(\bar{\vartheta})$, we get

$$\begin{aligned} \frac{1}{P_m} \sum_{k=1}^m p_k \psi(\vartheta_k) &= \frac{1}{P_m} \sum_{k=1}^m p_k \left[\psi(\rho_1) + (\vartheta_k - \rho_1)\psi'(\rho_1) \right. \\ & \quad \left. + \int_{\rho_1}^{\rho_2} G_1(\vartheta_k, x)\psi''(x) dx \right] \end{aligned} \quad (6)$$

and

$$\psi(\bar{\vartheta}) = \psi(\rho_1) + (\bar{\vartheta} - \rho_1)\psi'(\rho_1) + \int_{\rho_1}^{\rho_2} G_1(\bar{\vartheta}, x)\psi''(x) dx. \quad (7)$$

Subtracting (7) from (6), we get

$$\begin{aligned} \frac{1}{P_m} \sum_{k=1}^m p_k \psi(\vartheta_k) - \psi(\bar{\vartheta}) &= \psi(\rho_1) + \bar{\vartheta}\psi'(\rho_1) - \rho_1\psi'(\rho_1) \\ & \quad + \frac{1}{P_m} \sum_{k=1}^m p_k \int_{\rho_1}^{\rho_2} G_1(\vartheta_k, x)\psi''(x) dx \\ & \quad - \psi(\rho_1) - \bar{\vartheta}\psi'(\rho_1) + \rho_1\psi'(\rho_1) - \int_{\rho_1}^{\rho_2} G_1(\bar{\vartheta}, x)\psi''(x) dx \\ &= \int_{\rho_1}^{\rho_2} \left(\frac{1}{P_m} \sum_{k=1}^m p_k G_1(\vartheta_k, x) - G_1(\bar{\vartheta}, x) \right) \psi''(x) dx. \end{aligned} \quad (8)$$

Taking the absolute value of (8), we get

$$\begin{aligned} & \left| \frac{1}{P_m} \sum_{k=1}^m p_k \psi(\vartheta_k) - \psi(\bar{\vartheta}) \right| \\ &= \left| \int_{\rho_1}^{\rho_2} \left(\frac{1}{P_m} \sum_{k=1}^m p_k G_1(\vartheta_k, x) - G_1(\bar{\vartheta}, x) \right) \psi''(x) dx \right| \\ &\leq \int_{\rho_1}^{\rho_2} \left| \frac{1}{P_m} \sum_{k=1}^m p_k G_1(\vartheta_k, x) - G_1(\bar{\vartheta}, x) \right| |\psi''(x)| dx. \end{aligned} \quad (9)$$

Using the convexity of $G_1(z, x)$ and change of variable $x = t\rho_1 + (1-t)\rho_2$, $t \in [0, 1]$ in (9), we get

$$\begin{aligned} & \left| \frac{1}{P_m} \sum_{k=1}^m p_k \psi(\vartheta_k) - \psi(\bar{\vartheta}) \right| \\ &\leq (\rho_2 - \rho_1) \int_0^1 \left[\frac{1}{P_m} \sum_{k=1}^m p_k G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) - \right. \\ & \quad \left. G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) \right] |\psi''(t\rho_1 + (1-t)\rho_2)| dt. \end{aligned} \quad (10)$$

Since $|\psi''|$ is a strongly convex function with modulus c , therefore (10) becomes

$$\begin{aligned} & \left| \frac{1}{P_m} \sum_{k=1}^m p_k \psi(\vartheta_k) - \psi(\bar{\vartheta}) \right| \\ &\leq (\rho_2 - \rho_1) \int_0^1 \left[\frac{1}{P_m} \sum_{k=1}^m p_k G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) \right. \\ & \quad \left. - G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) \right] \\ & \quad \times (t|\psi''(\rho_1)| + (1-t)|\psi''(\rho_2)| - ct(1-t)(\rho_1 - \rho_2)^2) dt \\ &= (\rho_2 - \rho_1) \left[|\psi''(\rho_1)| \frac{1}{P_m} \sum_{k=1}^m p_k \int_0^1 t G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) dt \right. \\ & \quad + |\psi''(\rho_2)| \frac{1}{P_m} \sum_{k=1}^m p_k \int_0^1 G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) dt \\ & \quad \left. - |\psi''(\rho_2)| \frac{1}{P_m} \sum_{k=1}^m p_k \int_0^1 t G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt \right] \end{aligned}$$

$$\begin{aligned}
 & -|\psi''(\rho_1)| \int_0^1 t G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt \\
 & -|\psi''(\rho_2)| \int_0^1 G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt \\
 & +|\psi''(\rho_2)| \int_0^1 t G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt \\
 & -c(\rho_1-\rho_2)^2 \frac{1}{P} \sum_{k=1}^m p_k \int_0^1 t G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) dt \\
 & +c(\rho_1-\rho_2)^2 \frac{1}{P} \sum_{k=1}^m p_k \int_0^1 t^2 G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) dt \\
 & +c(\rho_1-\rho_2)^2 \int_0^1 t G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt \\
 & -c(\rho_1-\rho_2)^2 \int_0^1 t^2 G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt \Big]. \quad (11)
 \end{aligned}$$

Now utilizing change of the variable $x = t\rho_1 + (1-t)\rho_2$ for $t \in [0, 1]$, we have $t = \frac{\rho_2-x}{\rho_2-\rho_1}$ and $dt = -\frac{dx}{\rho_2-\rho_1}$. Also, when $t \rightarrow 0$ then $x \rightarrow \rho_2$ and when $t \rightarrow 1$ then $x \rightarrow \rho_1$. Using theses values as follows, we get

$$\begin{aligned}
 & \int_0^1 t^2 G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) dt \\
 & = \int_{\rho_1}^{\rho_2} \left(\frac{\rho_2-x}{\rho_2-\rho_1}\right)^2 G_1(\vartheta_k, x) \frac{dx}{\rho_2-\rho_1} \\
 & = \frac{1}{(\rho_2-\rho_1)^3} \int_{\rho_1}^{\rho_2} (\rho_2-x)^2 G_1(\vartheta_k, x) dx \\
 & = \frac{1}{(\rho_2-\rho_1)^3} \left[\int_{\rho_1}^{\vartheta_k} (\rho_2-x)^2 G_1(\vartheta_k, x) dx \right. \\
 & \quad \left. + \int_{\vartheta_k}^{\rho_2} (\rho_2-x)^2 G_1(\vartheta_k, x) dx \right]. \quad (12)
 \end{aligned}$$

Utilizing (3) in (12), we get

$$\begin{aligned}
 & \int_0^1 t^2 G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) dt \\
 & = \frac{1}{(\rho_2-\rho_1)^3} \left[\int_{\rho_1}^{\vartheta_k} (\rho_2-x)^2 (\rho_1-x) dx \right. \\
 & \quad \left. + \int_{\vartheta_k}^{\rho_2} (\rho_2-x)^2 (\rho_1-\vartheta_k) dx \right] \\
 & = \frac{1}{(\rho_2-\rho_1)^3} \left[\int_{\rho_1}^{\vartheta_k} (\rho_2^2 + x^2 - 2\rho_2 x) (\rho_1-x) dx \right. \\
 & \quad \left. + \int_{\vartheta_k}^{\rho_2} (\rho_2^2 + x^2 - 2\rho_2 x) (\rho_1-\vartheta_k) dx \right] \\
 & = \frac{1}{(\rho_2-\rho_1)^3} \left[\int_{\rho_1}^{\vartheta_k} (\rho_1\rho_2^2 - \rho_2^2 x + \rho_1 x^2 - x^3 - 2\rho_1\rho_2 x + 2\rho_2 x^2) dx \right. \\
 & \quad \left. + (\rho_1-\vartheta_k) \int_{\vartheta_k}^{\rho_2} (\rho_2^2 + x^2 - 2\rho_2 x) dx \right]
 \end{aligned}$$

$$\begin{aligned}
 & = \frac{1}{(\rho_2-\rho_1)^3} \left[\rho_1\rho_2^2(\vartheta_k-\rho_1) - \frac{\rho_2^2}{2}(\vartheta_k^2-\rho_1^2) + \frac{\rho_1}{3}(\vartheta_k^3-\rho_1^3) \right. \\
 & \quad \left. - \left(\frac{\vartheta_k^4}{4} - \frac{\rho_1^4}{4}\right) - \rho_1\rho_2(\vartheta_k^2-\rho_1^2) + \frac{2\rho_2}{3}(\vartheta_k^3-\rho_1^3) \right. \\
 & \quad \left. + (\rho_1-\vartheta_k) \left(\rho_2^2(\rho_2-\vartheta_k) + \frac{\rho_2^3}{3} - \frac{\vartheta_k^3}{3} - \rho_2(\rho_2^2-\vartheta_k^2) \right) \right] \\
 & = \frac{1}{(\rho_2-\rho_1)^3} \left[\rho_1\rho_2^2\vartheta_k - \rho_1^2\rho_2^2 - \frac{\rho_2^2\vartheta_k^2}{2} + \frac{\rho_1^2\rho_2^2}{2} + \frac{\rho_1\vartheta_k^3}{3} \right. \\
 & \quad \left. - \frac{\rho_1^4}{3} - \frac{\vartheta_k^4}{4} + \frac{\rho_1^4}{4} - \rho_1\rho_2\vartheta_k^2 + \rho_1^3\rho_2 + \frac{2\rho_2\vartheta_k^3}{3} - \frac{2\rho_2\rho_1^3}{3} \right. \\
 & \quad \left. + \rho_1\rho_2^3 - \rho_1\rho_2^2\vartheta_k + \frac{\rho_1\rho_2^3}{3} - \frac{\rho_1\vartheta_k^3}{3} - \rho_1\rho_2^3 + \rho_1\rho_2\vartheta_k^2 \right. \\
 & \quad \left. - \rho_2^3\vartheta_k + \rho_2^2\vartheta_k^2 - \frac{\rho_2^3\vartheta_k}{3} + \frac{\vartheta_k^4}{3} + \rho_2^3\vartheta_k - \rho_2\vartheta_k^3 \right] \\
 & = \frac{1}{(\rho_2-\rho_1)^3} \left(\frac{\rho_2^2\vartheta_k^2}{2} - \frac{\rho_1^2\rho_2^2}{2} - \frac{\rho_1^4}{12} + \frac{\vartheta_k^4}{12} + \frac{\rho_1^3\rho_2}{3} \right. \\
 & \quad \left. - \frac{\rho_2\vartheta_k^3}{3} + \frac{\rho_1\rho_2^3}{3} - \frac{\rho_2^3\vartheta_k}{3} \right). \quad (13)
 \end{aligned}$$

Replacing ϑ_k by $\bar{\vartheta}$ in (13), we get

$$\begin{aligned}
 & \int_0^1 t^2 G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt \\
 & = \frac{1}{(\rho_2-\rho_1)^3} \left(\frac{\rho_2^2(\bar{\vartheta})^2}{2} - \frac{\rho_1^2\rho_2^2}{2} - \frac{\rho_1^4}{12} + \frac{(\bar{\vartheta})^4}{12} \right. \\
 & \quad \left. + \frac{\rho_1^3\rho_2}{3} - \frac{\rho_2(\bar{\vartheta})^3}{3} + \frac{\rho_1\rho_2^3}{3} - \frac{\rho_2^3(\bar{\vartheta})}{3} \right). \quad (14)
 \end{aligned}$$

Similarly

$$\begin{aligned}
 & \int_0^1 t G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) dt \\
 & = \frac{1}{(\rho_2-\rho_1)^2} \left(\frac{\rho_2\vartheta_k^2}{2} - \frac{\rho_1^2\rho_2}{2} + \frac{\rho_1^3}{6} \right. \\
 & \quad \left. - \frac{\vartheta_k^3}{6} + \frac{\rho_1\rho_2^2}{2} - \frac{\rho_2^2\vartheta_k}{2} \right). \quad (15)
 \end{aligned}$$

Replacing ϑ_k by $\bar{\vartheta}$ in (15), we get

$$\begin{aligned}
 & \int_0^1 t G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt \\
 & = \frac{1}{(\rho_2-\rho_1)^2} \left(\frac{\rho_2(\bar{\vartheta})^2}{2} - \frac{\rho_1^2\rho_2}{2} + \frac{\rho_1^3}{6} \right. \\
 & \quad \left. - \frac{(\bar{\vartheta})^3}{6} + \frac{\rho_1\rho_2^2}{2} - \frac{\rho_2^2\bar{\vartheta}}{2} \right). \quad (16)
 \end{aligned}$$

Also,

$$\begin{aligned}
 & \int_0^1 G_1(\vartheta_k, t\rho_1 + (1-t)\rho_2) dt \\
 & = \frac{1}{\rho_2-\rho_1} \left(\frac{\vartheta_k^2}{2} - \frac{\rho_1^2}{2} + \rho_1\rho_2 - \rho_2\vartheta_k \right). \quad (17)
 \end{aligned}$$

Replacing ϑ_k by $\bar{\vartheta}$ in (17), we get

$$\int_0^1 G_1(\bar{\vartheta}, t\rho_1 + (1-t)\rho_2) dt = \frac{1}{\rho_2 - \rho_1} \left(\frac{(\bar{\vartheta})^2}{2} - \frac{\rho_1^2}{2} + \rho_1\rho_2 - \rho_2\bar{\vartheta} \right). \tag{18}$$

Substituting the values from (13)–(18) in (11), we get (5). \square

Remark 1 If we use the Green functions mentioned by equation numbers 1, 3, 4, and 5 in [21], instead of G_1 in Theorem 3, we obtain the same result (5).

As an application of Theorem 3, we derive a converse of the Hölder inequality.

Proposition 1 Let $[\rho_1, \rho_2]$ be a positive interval and $(a_1, \dots, a_m), (b_1, \dots, b_m)$ be two positive m -tuples, then (i) for $q > 1, p \in (1, 2) \cup (3, 4)$ with $1/p + 1/q = 1$ and $(\sum_{k=1}^m a_k b_k) / (\sum_{k=1}^m b_k^q), a_k b_k^{-q/p} \in [\rho_1, \rho_2]$ for $k = 1, \dots, m$, we have

$$\begin{aligned} & \left(\sum_{k=1}^m a_k^p \right)^{1/p} \left(\sum_{k=1}^m b_k^q \right)^{1/q} - \sum_{k=1}^m a_k b_k \\ & \leq \left\{ \frac{p(p-1)\rho_1}{4\rho_2^3(\rho_2-\rho_1)} \left(2\rho_2^4\rho_1^{p-3} - 2\rho_2^{p+1} + (p-2)(p-3)\rho_2^p(\rho_2-\rho_1) \right) \right. \\ & \quad \times \left(\frac{\sum_{k=1}^m a_k^2 b_k^{1-q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^2 \right) \\ & \quad + \frac{p(p-1)}{12\rho_2^3(\rho_2-\rho_1)} \left(2\rho_2^{p+1} - 2\rho_1^{p-2}\rho_2^3 - (p-2)(p-3)\rho_2^{p-1}(\rho_2-\rho_1^2) \right) \\ & \quad \times \left(\frac{\sum_{k=1}^m a_k^3 b_k^{1-2q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^3 \right) \\ & \quad + \frac{p(p-1)(p-2)(p-3)\rho_2^{p-4}}{24} \\ & \quad \left. \times \left(\frac{\sum_{k=1}^m a_k^4 b_k^{1-3q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^4 \right) \right\}^{1/p} \sum_{k=1}^m b_k^q. \tag{19} \end{aligned}$$

(ii) For $q > 1, p > 4$ with $1/p + 1/q = 1$ and $(\sum_{k=1}^m a_k b_k) / (\sum_{k=1}^m b_k^q), a_k b_k^{-q/p} \in [\rho_1, \rho_2]$ for $k =$

$1, \dots, m$, we have

$$\begin{aligned} & \left(\sum_{k=1}^m a_k^p \right)^{1/p} \left(\sum_{k=1}^m b_k^q \right)^{1/q} - \sum_{k=1}^m a_k b_k \\ & \leq \left\{ \frac{p(p-1)\rho_2}{4\rho_1^3(\rho_2-\rho_1)} \left(2\rho_1^{p+1} - 2\rho_1^4\rho_2^{p-3} + (p-2)(p-3)\rho_1^p(\rho_2-\rho_1) \right) \right. \\ & \quad \times \left(\frac{\sum_{k=1}^m a_k^2 b_k^{1-q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^2 \right) \\ & \quad + \frac{p(p-1)}{12\rho_1^3(\rho_2-\rho_1)} \left(2\rho_1^3\rho_2^{p-2} - 2\rho_1^{p+1} - (p-2)(p-3)\rho_1^{p-1}(\rho_2-\rho_1^2) \right) \\ & \quad \times \left(\frac{\sum_{k=1}^m a_k^3 b_k^{1-2q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^3 \right) \\ & \quad + \frac{p(p-1)(p-2)(p-3)\rho_1^{p-4}}{24} \\ & \quad \left. \times \left(\frac{\sum_{k=1}^m a_k^4 b_k^{1-3q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^4 \right) \right\}^{1/p} \sum_{k=1}^m b_k^q. \tag{20} \end{aligned}$$

Proof: (i) Let $\psi(x) = x^p, x \in [\rho_1, \rho_2]$, then $\psi''(x) = p(p-1)x^{p-2} > 0$, which shows that the function ψ is convex. Also, $\psi''''(x) = |\psi''''(x)| = p(p-1)(p-2)(p-3)x^{p-4}$, which is a decreasing function for given values of p , and $\psi''''(x) \geq 2 \left(\frac{p(p-1)(p-2)(p-3)\rho_2^{p-4}}{2} \right)$. Therefore $|\psi''|$ is strongly convex function with modulus $c = \frac{p(p-1)(p-2)(p-3)\rho_2^{p-4}}{2}$, hence using (5) for $\psi(x) = x^p, p_k = b_k^q$ and $\vartheta_k = a_k b_k^{-q/p}$, we derive

$$\begin{aligned} & \left(\left(\sum_{k=1}^m a_k^p \right) \left(\sum_{k=1}^m b_k^q \right)^{p-1} - \left(\sum_{k=1}^m a_k b_k \right)^p \right)^{1/p} \\ & \leq \left\{ \frac{p(p-1)\rho_1}{4\rho_2^3(\rho_2-\rho_1)} \left(2\rho_2^4\rho_1^{p-3} - 2\rho_2^{p+1} + (p-2)(p-3)\rho_2^p(\rho_2-\rho_1) \right) \right. \\ & \quad \times \left(\frac{\sum_{k=1}^m a_k^2 b_k^{1-q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^2 \right) \\ & \quad + \frac{p(p-1)}{12\rho_2^3(\rho_2-\rho_1)} \left(2\rho_2^{p+1} - 2\rho_1^{p-2}\rho_2^3 - (p-2)(p-3)\rho_2^{p-1}(\rho_2-\rho_1^2) \right) \\ & \quad \times \left(\frac{\sum_{k=1}^m a_k^3 b_k^{1-2q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^3 \right) \\ & \quad + \frac{p(p-1)(p-2)(p-3)\rho_2^{p-4}}{24} \\ & \quad \left. \times \left(\frac{\sum_{k=1}^m a_k^4 b_k^{1-3q/p}}{\sum_{k=1}^m b_k^q} - \left[\frac{\sum_{k=1}^m a_k b_k}{\sum_{k=1}^m b_k^q} \right]^4 \right) \right\}^{1/p} \sum_{k=1}^m b_k^q. \tag{21} \end{aligned}$$

Utilizing the inequality $x^\ell - y^\ell \leq (x-y)^\ell, 0 \leq y \leq x, \ell \in [0, 1]$ for $x = \left(\sum_{k=1}^m a_k^p \right) \left(\sum_{k=1}^m b_k^q \right)^{p-1}, y = \left(\sum_{k=1}^m a_k b_k \right)^p$ and $\ell = 1/p$, we obtain

$$\begin{aligned} & \left(\sum_{k=1}^m a_k^p \right)^{1/p} \left(\sum_{k=1}^m b_k^q \right)^{1/q} - \sum_{k=1}^m a_k b_k \\ & \leq \left(\left(\sum_{k=1}^m a_k^p \right) \left(\sum_{k=1}^m b_k^q \right)^{p-1} - \left(\sum_{k=1}^m a_k b_k \right)^p \right)^{1/p}. \tag{22} \end{aligned}$$

From (21) and (22), we get (19).

(ii) Adopting the procedure of part (i), where $\psi'''' = |\psi''|''$ becomes an increasing function for given values of p and thus acquiring $c = \frac{p(p-1)(p-2)(p-3)\rho_1^{p-4}}{2}$ and we conclude the result (20). \square

The following theorem is the integral version of Theorem 3. This provides an improvement of the bound for Jensen's gap given in [17].

Theorem 4 Let $\psi \in C^2[\rho_1, \rho_2]$ be such that $|\psi''|$ is strongly convex function with modulus c . Let $\eta_1, \eta_2 : [\gamma_1, \gamma_2] \rightarrow \mathbb{R}$ be two integrable functions such that $\eta_1(y) \in [\rho_1, \rho_2]$ and $\eta_2(y) \geq 0$ for all $y \in [\gamma_1, \gamma_2]$ with $D := \int_{\gamma_1}^{\gamma_2} \eta_2(y) dy > 0$, $\bar{\eta} = \frac{1}{D} \int_{\gamma_1}^{\gamma_2} \eta_1(y)\eta_2(y) dy$. Then we have

$$\begin{aligned} & \left| \frac{1}{D} \int_{\gamma_1}^{\gamma_2} (\psi \circ \eta_1)(y)\eta_2(y) dy - \psi(\bar{\eta}) \right| \\ & \leq \frac{\rho_2|\psi''(\rho_1)| - \rho_1|\psi''(\rho_2)| + c\rho_1\rho_2(\rho_2 - \rho_1)}{2(\rho_2 - \rho_1)} \\ & \quad \times \left(\frac{1}{D} \int_{\gamma_1}^{\gamma_2} \eta_1^2(y)\eta_2(y) dy - (\bar{\eta})^2 \right) \\ & + \frac{|\psi''(\rho_2)| - |\psi''(\rho_1)| - c(\rho_2^2 - \rho_1^2)}{6(\rho_2 - \rho_1)} \\ & \quad \times \left(\frac{1}{D} \int_{\gamma_1}^{\gamma_2} \eta_1^3(y)\eta_2(y) dy - (\bar{\eta})^3 \right) \\ & + \frac{c}{12} \left(\frac{1}{D} \int_{\gamma_1}^{\gamma_2} \eta_1^4(y)\eta_2(y) dy - (\bar{\eta})^4 \right). \end{aligned} \tag{23}$$

Proof: The proof is analogous to the proof of Theorem 3. \square

As an application of Theorem 4, we give a new bound for the Hermite-Hadamard gap.

Corollary 1 Let $\phi \in C^2[\gamma_1, \gamma_2]$ be a function such that $|\phi''|$ is strongly convex function with modulus c , then

$$\begin{aligned} & \left| \frac{1}{\gamma_2 - \gamma_1} \int_{\gamma_1}^{\gamma_2} \phi(y) dy - \phi\left(\frac{\gamma_1 + \gamma_2}{2}\right) \right| \\ & \leq \frac{(|\phi''(\gamma_1)| + |\phi''(\gamma_2)|)(\gamma_2 - \gamma_1)^2}{48} \\ & + \frac{3c}{320} (4\gamma_1\gamma_2^3 + 4\gamma_1^3\gamma_2 - 6\gamma_1^2\gamma_2^2 - \gamma_1^4 - \gamma_2^4). \end{aligned} \tag{24}$$

Proof: Using (23) for $\psi = \phi$, $[\rho_1, \rho_2] = [\gamma_1, \gamma_2]$, $\eta_2(y) = 1$ and $\eta_1(y) = y$ for all $y \in [\gamma_1, \gamma_2]$, we get (24). \square

We demonstrate the following two numerical experiments to show that the newly obtained bounds for Jensen's gap are some improvements of the earlier bounds given in [17, 18]. This leads to the conclusion that for better results strongly convex functions may be preferred on convex functions.

Example 1 Let $\psi(y) = y^4$, $\eta_1(y) = y$, $\eta_2(y) = 1$ for all $y \in [0, 1]$ then $\psi''(y) = 12y^2 > 0$, $\psi''''(y) = |\psi''|''(y) = 24 = 2(12)$ for all $y \in [0, 1]$. Which shows that ψ is convex function and $|\psi''|$ is strongly convex function on $[0, 1]$ with modulus $c = 12$. Also, $\eta_1(y) \in [0, 1]$ for all $y \in [0, 1]$, therefore using inequality (23) for these facts with $[\rho_1, \rho_2] = [\gamma_1, \gamma_2] = [0, 1]$, we obtain $\int_0^1 \psi(\eta_1(y)) dy - \psi(\int_0^1 \eta_1(y) dy) = 0.2 - 0.0625 = 0.1375$ and also its corresponding right hand side gives 0.1375. Thus from inequality (23) we conclude that

$$0.1375 \leq 0.1375. \tag{25}$$

Now with aforementioned parameters of this example, but considering $|\psi''|$ as a convex function, the estimate calculated in [17] for the Jensen gap 0.1375 is 0.25, i.e.

$$0.1375 < 0.25. \tag{26}$$

From (25) and (26) we conclude that using strongly convex functions gives better estimates for the Jensen gap instead of using convex functions.

Example 2 Let $\psi(y) = e^y$, $\eta_1(y) = y^2$, $\eta_2(y) = 1$ for all $y \in [0, 1]$ then $\psi''(y) = e^y > 0$, $\psi''''(y) = |\psi''|''(y) = e^y \geq 2(e^0/2) = 2(0.5)$ for all $y \in [0, 1]$. Which shows that ψ is a convex function and the function $|\psi''|$ is strongly convex with modulus $c = 0.5$. Also, $\eta_1(y) \in [0, 1]$ for all $y \in [0, 1]$, therefore using inequality (23) for the above facts with $[\rho_1, \rho_2] = [\gamma_1, \gamma_2] = [0, 1]$, we obtain $\int_0^1 \psi(\eta_1(y)) dy - \psi(\int_0^1 \eta_1(y) dy) = 1.4627 - 1.3956 = 0.0671$ and right hand side of inequality (23) gives 0.0701. Thus from inequality (23), we deduce the following result

$$0.0671 < 0.0701. \tag{27}$$

Now for aforementioned parameters, but instead providing $|\psi''|$ as a convex function, the bound for the Jensen gap 0.0671 is 0.0748 [17], i.e.

$$0.0671 < 0.0748. \tag{28}$$

From (27) and (28) we conclude that using strongly convex functions gives better estimates for the Jensen gap instead of using convex functions.

APPLICATIONS IN INFORMATION THEORY

Information theory is a useful mathematical tool, not limited to communication, but more technically as an important part of probability theory. It emerged from Shannon [22] by considering stochastic process as a source of information. Information theory has deep connections with diverse topics as statistical inference, artificial intelligence, statistical mechanics and biological evolution. For a probability space $U(S, B(S), U)$, the information content $I(E)$ of an event $E \in B(S)$

is defined to be negative log of $U(E)$, that is $I(E) = -\log_b U(E)$. The base of log represent the unit of information. The key concept of information theory is entropy, the measure of uncertainty of random variable. For a discrete random variable X with possible outcomes $\{x_1, \dots, x_m\}$ with corresponding probabilities $\{w_1, \dots, w_m\}$, and information content $I(X) = \{I(w_1), \dots, I(w_m)\}$, the Shannon entropy, $Z_s(\mathbf{w})$ is the expected value of $I(X)$.

$$Z_s(\mathbf{w}) = E(I(X)) = \sum_{k=1}^m w_k \log \frac{1}{w_k}, \quad w_k \neq 0 \quad (29)$$

with the convention that $Z_s(\mathbf{w}) = 0$, when $w_k = 0$ for all $k \in \{1, \dots, m\}$. $Z_s(\mathbf{w})$ is the only function satisfying three natural properties: (i) $Z_s(\mathbf{w})$ is positive or null; (ii) uniform distribution maximizes $Z_s(\mathbf{w})$; and (iii) $Z_s(\mathbf{w})$ has additive property of successive information.

A commonly used information criteria to determine model discrepancy is Kullback-Leibler divergence or KL-divergence. It is a non-symmetric measure of variation between two probability distributions, one being the fitted model and the other being the reference model. Let r_k and w_k for $k = 1, \dots, m$, be two probability distributions on the same random variable X . The KL-divergence is the expectation of the log ratio of r_k and w_k ,

$$Z_{kl}(\mathbf{r}, \mathbf{w}) = \sum_{k=1}^m r_k \log \frac{r_k}{w_k},$$

where $Z_{kl}(\mathbf{r}, \mathbf{w}) = 0$, implies that the two distributions are identical. The KL-divergence is a special case of Csiszár f -divergence with diverse applications in applied statistics, fluid mechanics, neuroscience and machine learning.

Information theory is also involved in analysis of human language. Words in a human language occur systematically in such a way that very few frequently used words account for most of the tokens in the text. The distribution of words in human corpus roughly follow a power law known as Zipf's law. Zipf [23] observed that if words are ranked according to their frequency f , in decreasing order, frequency is a non-linear decreasing function of rank k , that is, $f_k = C/k^s$ with positive parameters, C and s , to be estimated from given data. Due to lack of fit in the low and high rank regions, Mandelbrot [24] generalized Zipf's law by adding a nonnegative parameter θ as follows,

$$f_k = \frac{C}{(k + \theta)^s},$$

which tends to Zipf's law for $\theta = 0$. The probability mass function for the k -th word in corpus of m words is:

$$f(k, m, \theta, s) = \frac{1}{(k + \theta)^s M_{m, \theta, s}},$$

for $k = 1, \dots, m$. $M_{m, \theta, s} = \sum_{k=1}^m \frac{1}{(k + \theta)^s}$ is generalized harmonic number. Apart from linguistics, Zipf distribution is used in city population, web site traffic, earthquake magnitude and in economics it gives best fit to affluent people in the community. In connection to the attitude of information theory, entropies are utilized to compute the amount of information in written text. The Zipf-Mandelbrot entropy mentioned in [6] is given by:

$$Z(M, \theta, s) = \frac{s}{M_{m, \theta, s}} \sum_{k=1}^m \frac{\log(k + \theta)}{(k + \theta)^s} + \log M_{m, \theta, s}.$$

Results for Csiszár divergence and its special cases

This subsection addresses the improvements of some earlier estimates given for various divergences in [17, 18].

Definition 2 [Csiszár divergence] Let $f : [\rho_1, \rho_2] \subset \mathbb{R} \rightarrow \mathbb{R}$ be a function, then Csiszár divergence for $\mathbf{r} = (r_1, \dots, r_m) \in \mathbb{R}^n$ and $\mathbf{w} = (w_1, \dots, w_m) \in \mathbb{R}_+^n$ with $r_k/w_k \in [\rho_1, \rho_2]$ ($k = 1, \dots, m$), is defined as

$$Z_c(\mathbf{r}, \mathbf{w}) = \sum_{k=1}^m w_k f\left(\frac{r_k}{w_k}\right).$$

Theorem 5 Let $f \in C^2[\rho_1, \rho_2]$ be a function such that the function $|f''|$ is strongly convex with modulus c . Let $\mathbf{r} = (r_1, \dots, r_m) \in \mathbb{R}^n$ and $\mathbf{w} = (w_1, \dots, w_m) \in \mathbb{R}_+^n$ be such that $(\sum_{k=1}^m r_k)(\sum_{k=1}^m w_k)$, $r_k/w_k \in [\rho_1, \rho_2]$ for $k = 1, \dots, m$, then

$$\begin{aligned} & \left| \frac{1}{\sum_{k=1}^m w_k} Z_c(\mathbf{r}, \mathbf{w}) - f\left(\frac{\sum_{k=1}^m r_k}{\sum_{k=1}^m w_k}\right) \right| \\ & \leq \frac{\rho_2 |\psi''(\rho_1)| - \rho_1 |\psi''(\rho_2)| + c \rho_1 \rho_2 (\rho_2 - \rho_1)}{2(\rho_2 - \rho_1)} \\ & \quad \times \left(\frac{1}{\sum_{k=1}^m w_k} \sum_{k=1}^m \frac{r_k^2}{w_k} - \left[\frac{\sum_{k=1}^m r_k}{\sum_{k=1}^m w_k} \right]^2 \right) \\ & \quad + \frac{|\psi''(\rho_2)| - |\psi''(\rho_1)| - c(\rho_2^2 - \rho_1^2)}{6(\rho_2 - \rho_1)} \\ & \quad \times \left(\frac{1}{\sum_{k=1}^m w_k} \sum_{k=1}^m \frac{r_k^3}{w_k^2} - \left[\frac{\sum_{k=1}^m r_k}{\sum_{k=1}^m w_k} \right]^3 \right) \\ & \quad + \frac{c}{12} \left(\frac{1}{\sum_{k=1}^m w_k} \sum_{k=1}^m \frac{r_k^4}{w_k^3} - \left[\frac{\sum_{k=1}^m r_k}{\sum_{k=1}^m w_k} \right]^4 \right). \quad (30) \end{aligned}$$

Proof: Utilizing $\psi = f$, $\vartheta_k = r_k/w_k$ and $p_k = w_k/\sum_{k=1}^m w_k$ in (5), we obtain (30). \square

Definition 3 [Rényi-divergence] For a nonnegative real number $\tau \neq 1$ and two positive probability distributions $\mathbf{r} = (r_1, \dots, r_m)$ and $\mathbf{w} = (w_1, \dots, w_m)$, the Rényi-divergence is defined as

$$Z_{re}(\mathbf{r}, \mathbf{w}) = \frac{1}{\tau - 1} \log \left(\sum_{k=1}^m r_k^\tau w_k^{1-\tau} \right).$$

Corollary 2 Let $[\rho_1, \rho_2]$ be a positive interval and $\mathbf{r} = (r_1, \dots, r_m)$, $\mathbf{w} = (w_1, \dots, w_m)$ be positive probability distributions with $\sum_{k=1}^m w_k (r_k/w_k)^\tau$, $(r_k/w_k)^{\tau-1} \in [\rho_1, \rho_2]$ for $k = 1, \dots, m$, provided $\tau > 1$, then

$$\begin{aligned} Z_{re}(\mathbf{r}, \mathbf{w}) &= \frac{1}{\tau-1} \sum_{k=1}^m r_k \log \left(\frac{r_k}{w_k} \right)^{\tau-1} \\ &\leq \frac{\rho_2^5 + 2\rho_1^3 \rho_2^3 - 3\rho_1^4 \rho_2}{2(\tau-1)\rho_1^2 \rho_2^4 (\rho_2 - \rho_1)} \\ &\quad \times \left(\sum_{k=1}^m r_k \left(\frac{r_k}{w_k} \right)^{2(\tau-1)} - \left(\sum_{k=1}^m r_k^\tau w_k^{1-\tau} \right)^2 \right) \\ &\quad + \frac{3\rho_1^4 - \rho_2^4 - 2\rho_1^2 \rho_2^2}{6(\tau-1)\rho_1^2 \rho_2^4 (\rho_2 - \rho_1)} \\ &\quad \times \left(\sum_{k=1}^m r_k \left(\frac{r_k}{w_k} \right)^{3(\tau-1)} - \left(\sum_{k=1}^m r_k^\tau w_k^{1-\tau} \right)^3 \right) \\ &\quad + \frac{1}{4(\tau-1)\rho_2^4} \left(\sum_{k=1}^m r_k \left(\frac{r_k}{w_k} \right)^{4(\tau-1)} - \left(\sum_{k=1}^m r_k^\tau w_k^{1-\tau} \right)^4 \right). \end{aligned} \quad (31)$$

Proof: Let $\psi(x) = -(\log x)/(\tau-1)$, $x \in [\rho_1, \rho_2]$, then $\psi''(x) = 1/(\tau-1)x^2 > 0$, which shows that the function ψ is convex. Also, $\psi''''(x) = |\psi''|''(x) = 6/(\tau-1)x^4 \geq 2(3/(\tau-1)\rho_2^4)$, which implies that the function $|\psi''|$ is strongly convex with modulus $c = 3/(\tau-1)\rho_2^4$, therefore we obtain (31) by using (5) for $\psi(x) = -(\log x)/(\tau-1)$, $p_k = r_k$ and $\vartheta_k = (r_k/w_k)^{\tau-1}$. \square

In the following corollary, we present an improved estimate for Shannon entropy in discrete form.

Corollary 3 Let $\mathbf{w} = (w_1, \dots, w_m)$ be a positive probability distribution with $1/w_k \in [\rho_1, \rho_2] \subseteq \mathbb{R}^+$ for $k = 1, \dots, m$, then we have

$$\begin{aligned} \log m - Z_s(\mathbf{w}) &\leq \frac{\rho_2^2(\rho_2^2 + \rho_1 \rho_2 + \rho_1^2) + 3\rho_1^3 \rho_2}{2\rho_1^2 \rho_2^4} \left(\sum_{k=1}^m \frac{1}{w_k} - m^2 \right) \\ &\quad - \frac{(\rho_1 + \rho_2)(\rho_2^2 + 3\rho_1^2)}{6\rho_1^2 \rho_2^4} \left(\sum_{k=1}^m \frac{1}{w_k^2} - m^3 \right) \\ &\quad + \frac{1}{4\rho_2^4} \left(\sum_{k=1}^m \frac{1}{w_k^3} - m^4 \right). \end{aligned} \quad (32)$$

Proof: Let $f(x) = -\log x$, $x \in [\rho_1, \rho_2]$, then $f''(x) = 1/x^2 > 0$, which shows that the function f is convex. Also, we know from Theorem 2 that a function f is strongly convex with modulus c if $f''(x) \geq 2c$ for all x in its domain. Now, here we have $f''''(x) = |f''|''(x) = 6/x^4 \geq 2(3/\rho_2^4)$, which shows that $|f''|$ is a strongly convex function with $c = 3/\rho_2^4 > 0$. Therefore, using (30) for $f(x) = -\log x$, $(r_1, \dots, r_m) = (1, \dots, 1)$ and $c = 3/\rho_2^4$, we obtain (32). \square

The integral form of (29), for a positive probability density function $\xi(y)$ defined on $[\gamma_1, \gamma_2]$ is given by

$$Z_s(\xi) = - \int_{\gamma_1}^{\gamma_2} \xi(y) \log \xi(y) dy.$$

Following is the integral version of Corollary 3.

Corollary 4 Let $\xi(y)$ be a positive probability density function defined on $[\gamma_1, \gamma_2]$ with $1/\xi(y) \in [\rho_1, \rho_2] \subseteq \mathbb{R}^+$ for all $y \in [\gamma_1, \gamma_2]$, then

$$\begin{aligned} \log(\gamma_2 - \gamma_1) - Z_s(\xi) &\leq \frac{\rho_2^2(\rho_2^2 + \rho_1 \rho_2 + \rho_1^2) + 3\rho_1^3 \rho_2}{2\rho_1^2 \rho_2^4} \left(\int_{\gamma_1}^{\gamma_2} \frac{1}{\xi(y)} dy - (\gamma_2 - \gamma_1)^2 \right) \\ &\quad - \frac{(\rho_1 + \rho_2)(\rho_2^2 + 3\rho_1^2)}{6\rho_1^2 \rho_2^4} \left(\int_{\gamma_1}^{\gamma_2} \frac{1}{\xi^2(y)} dy - (\gamma_2 - \gamma_1)^3 \right) \\ &\quad + \frac{1}{4\rho_2^4} \left(\int_{\gamma_1}^{\gamma_2} \frac{1}{\xi^3(y)} dy - (\gamma_2 - \gamma_1)^4 \right). \end{aligned} \quad (33)$$

Proof: The proof is analogous to the proof of Corollary 3. \square

Example 3 Let $\xi(y) = 1/y \ln 2$ for all $y \in [1, 2] = [\gamma_1, \gamma_2]$ be a reciprocal distribution, then $1/\xi(y) \in [1/2, 3/2] = [\rho_1, \rho_2]$ for all $y \in [1, 2]$. Using (33) for these values we obtain

$$0.0199 < 0.0401. \quad (34)$$

Now for the given values, the corresponding estimate in [17] has been calculated as

$$0.0199 < 0.0429. \quad (35)$$

From (34) and (35), we conclude that the bound given by (33) is better than the bound provided in [17].

Corollary 5 If $\mathbf{r} = (r_1, \dots, r_m)$ and $\mathbf{w} = (w_1, \dots, w_m)$ are positive probability distributions with $r_k/w_k \in [\rho_1, \rho_2] \subseteq \mathbb{R}^+$ for $k = 1, \dots, m$, then

$$\begin{aligned} Z_{kl}(\mathbf{r}, \mathbf{w}) &\leq \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{2\rho_1 \rho_2^2} \left(\sum_{k=1}^m \frac{r_k^2}{w_k} - 1 \right) \\ &\quad - \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{6\rho_1 \rho_2^3} \left(\sum_{k=1}^m \frac{r_k^3}{w_k^2} - 1 \right) + \frac{1}{12\rho_2^3} \left(\sum_{k=1}^m \frac{r_k^4}{w_k^3} - 1 \right). \end{aligned} \quad (36)$$

Proof: Let $f(x) = x \log x$, $x \in [\rho_1, \rho_2]$, then $f''(x) = 1/x > 0$, which shows that the function f is convex. Also, $f''''(x) = |f''|''(x) = 2/x^3 \geq 2(1/\rho_2^3)$, which presents that the function $|f''|$ is strongly convex with modulus $c = 1/\rho_2^3 > 0$. Thus we get (36) by using (30) for $f(x) = x \log x$. \square

Definition 4 [Bhattacharyya coefficient]

Bhattacharyya coefficient for two positive probability distributions $\mathbf{r} = (r_1, \dots, r_m)$ and $\mathbf{w} = (w_1, \dots, w_m)$ is defined by

$$Z_b(\mathbf{r}, \mathbf{w}) = \sum_{k=1}^m \sqrt{r_k w_k}.$$

Corollary 6 Let $\mathbf{r} = (r_1, \dots, r_m)$ and $\mathbf{w} = (w_1, \dots, w_m)$ be positive probability distributions such that $r_k/w_k \in [\rho_1, \rho_2] \subseteq \mathbb{R}^+$ for $k = 1, \dots, m$, then

$$1 - Z_b(\mathbf{r}, \mathbf{w}) \leq \frac{8\rho_2(\rho_2^{\frac{5}{2}} - \rho_1^{\frac{5}{2}}) + 15\rho_1^{\frac{5}{2}}(\rho_2 - \rho_1)}{64\rho_1^{\frac{3}{2}}\rho_2^{\frac{5}{2}}(\rho_2 - \rho_1)} \left(\sum_{k=1}^m \frac{r_k^2}{w_k} - 1 \right) - \frac{8\rho_2^2(\rho_2^{\frac{3}{2}} - \rho_1^{\frac{3}{2}}) + 15\rho_1^{\frac{3}{2}}(\rho_2^2 - \rho_1^2)}{192\rho_1^{\frac{3}{2}}\rho_2^{\frac{7}{2}}(\rho_2 - \rho_1)} \left(\sum_{k=1}^m \frac{r_k^3}{w_k^2} - 1 \right) + \frac{5}{128\rho_2^{\frac{7}{2}}} \left(\sum_{k=1}^m \frac{r_k^4}{w_k^3} - 1 \right). \quad (37)$$

Proof: Let $f(x) = -\sqrt{x}$ for $x \in [\rho_1, \rho_2]$, then $f''(x) = 1/4x^{3/2} > 0$ and $f'''(x) = |f''''(x)| = 15/16x^{7/2} \geq 2(15/32\rho_2^{7/2})$. Which presents that f is convex function while the function $|f''|$ is strongly convex with modulus $c = 15/32\rho_2^{7/2}$. Therefore, we obtain (37) by using (30) for $f(x) = -\sqrt{x}$. \square

Definition 5 [Hellinger distance] For two positive probability distributions $\mathbf{r} = (r_1, \dots, r_m)$, $\mathbf{w} = (w_1, \dots, w_m)$ the Hellinger distance is defined as

$$Z_h(\mathbf{r}, \mathbf{w}) = \frac{1}{2} \sum_{k=1}^m (\sqrt{r_k} - \sqrt{w_k})^2.$$

Corollary 7 If $[\rho_1, \rho_2]$ is a positive interval and $\mathbf{r} = (r_1, \dots, r_m)$, $\mathbf{w} = (w_1, \dots, w_m)$ are positive probability distributions such that $r_k/w_k \in [\rho_1, \rho_2]$ for $k = 1, \dots, m$, then

$$Z_h(\mathbf{r}, \mathbf{w}) \leq \frac{8\rho_2(\rho_2^{\frac{5}{2}} - \rho_1^{\frac{5}{2}}) + 15\rho_1^{\frac{5}{2}}(\rho_2 - \rho_1)}{64\rho_1^{\frac{3}{2}}\rho_2^{\frac{5}{2}}(\rho_2 - \rho_1)} \left(\sum_{k=1}^m \frac{r_k^2}{w_k} - 1 \right) - \frac{8\rho_2^2(\rho_2^{\frac{3}{2}} - \rho_1^{\frac{3}{2}}) + 15\rho_1^{\frac{3}{2}}(\rho_2^2 - \rho_1^2)}{192\rho_1^{\frac{3}{2}}\rho_2^{\frac{7}{2}}(\rho_2 - \rho_1)} \left(\sum_{k=1}^m \frac{r_k^3}{w_k^2} - 1 \right) + \frac{5}{128\rho_2^{\frac{7}{2}}} \left(\sum_{k=1}^m \frac{r_k^4}{w_k^3} - 1 \right). \quad (38)$$

Proof: Let $f(x) = (1 - \sqrt{x})^2/2$, $x \in [\rho_1, \rho_2]$, then $f''(x) = 1/4x^{3/2} > 0$ and $f'''(x) = |f''''(x)| = 15/16x^{7/2} \geq 2(15/32\rho_2^{7/2})$. This shows that the function f is convex and the function $|f''|$ is strongly convex with modulus $c = 15/32\rho_2^{7/2}$. Hence using (30) for $f(x) = (1 - \sqrt{x})^2/2$, we obtain (38). \square

Definition 6 [Triangular discrimination] For two positive probability distributions $\mathbf{r} = (r_1, \dots, r_m)$, $\mathbf{w} = (w_1, \dots, w_m)$, the triangular discrimination is defined as

$$Z_{\Delta}(\mathbf{r}, \mathbf{w}) = \sum_{k=1}^m \frac{(r_k - w_k)^2}{r_k + w_k}.$$

Corollary 8 Let $[\rho_1, \rho_2] \subseteq \mathbb{R}^+$ and $\mathbf{r} = (r_1, \dots, r_m)$, $\mathbf{w} = (w_1, \dots, w_m)$ be positive probability distributions with $r_k/w_k \in [\rho_1, \rho_2]$ for $k = 1, \dots, m$, then

$$Z_{\Delta}(\mathbf{r}, \mathbf{w}) \leq \frac{8\rho_2(\rho_2+1)^5 - 8\rho_1(\rho_1+1)^3(\rho_2+1)^2 + 48\rho_1\rho_2(\rho_2-\rho_1)(\rho_1+1)^3}{2(\rho_1+1)^3(\rho_2+1)^5(\rho_2-\rho_1)} \times \left(\sum_{k=1}^m \frac{r_k^2}{w_k} - 1 \right) + \frac{8(\rho_1+1)^3(\rho_2+1)^2 - 8(\rho_2+1)^5 - 48(\rho_2^2 - \rho_1^2)(\rho_1+1)^3}{6(\rho_1+1)^3(\rho_2+1)^5(\rho_2-\rho_1)} \times \left(\sum_{k=1}^m \frac{r_k^3}{w_k^2} - 1 \right) + \frac{4}{(\rho_2+1)^5} \left(\sum_{k=1}^m \frac{r_k^4}{w_k^3} - 1 \right). \quad (39)$$

Proof: Let $f(x) = (x - 1)^2/(x + 1)$ for $x \in [\rho_1, \rho_2]$, then $f''(x) = 8/(x + 1)^3 > 0$ and $f'''(x) = |f''''(x)| = 96/(x + 1)^5 \geq 2(48/(\rho_2 + 1)^5)$. This shows that f is a convex function while $|f''|$ is a strongly convex function with modulus $c = 48/(\rho_2 + 1)^5$. Thus using (30) for $f(x) = (x - 1)^2/(x + 1)$, we get (39). \square

Results for the Zipf-Mandelbrot entropy

This subsection is devoted to some new results around the Zipf-Mandelbrot entropy.

Corollary 9 Let $\theta \geq 0$, $s > 0$ and $w_k \geq 0$ for $k = 1, \dots, m$ with $\sum_{k=1}^m w_k = 1$ and $1/w_k(k + \theta)^s M_{m,\theta,s} \in [\rho_1, \rho_2] \subseteq \mathbb{R}^+$ ($k = 1, \dots, m$), then

$$-Z(M, \theta, s) - \frac{1}{M_{m,\theta,s}} \sum_{k=1}^m \frac{\log w_k}{(k + \theta)^s} \leq \frac{\rho_1^2 + \rho_1\rho_2 + \rho_2^2}{2\rho_1\rho_2^2} \left(\sum_{k=1}^m \frac{1}{w_k(k + \theta)^{2s} M_{m,\theta,s}^2} - 1 \right) - \frac{\rho_1^2 + \rho_1\rho_2 + \rho_2^2}{6\rho_1\rho_2^3} \left(\sum_{k=1}^m \frac{1}{w_k^2(k + \theta)^{3s} M_{m,\theta,s}^3} - 1 \right) + \frac{1}{12\rho_2^3} \left(\sum_{k=1}^m \frac{1}{w_k^3(k + \theta)^{4s} M_{m,\theta,s}^4} - 1 \right). \quad (40)$$

Proof: For $r_k = 1/(k + \theta)^s M_{m,\theta,s}$, $k = 1, \dots, m$, we have

$$\begin{aligned} & \sum_{k=1}^m r_k \log \frac{r_k}{w_k} \\ &= \sum_{k=1}^m \frac{1}{(k + \theta)^s M_{m,\theta,s}} (-s \log(k + \theta) - \log M_{m,\theta,s} - \log w_k) \\ &= -Z(M, \theta, s) - \frac{1}{M_{m,\theta,s}} \sum_{k=1}^m \frac{\log w_k}{(k + \theta)^s}. \end{aligned} \quad (41)$$

Also,

$$\begin{aligned} & \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{2\rho_1 \rho_2^2} \left(\sum_{k=1}^m \frac{r_k^2}{w_k} - 1 \right) \\ & - \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{6\rho_1 \rho_2^3} \left(\sum_{k=1}^m \frac{r_k^3}{w_k^2} - 1 \right) + \frac{1}{12\rho_2^3} \left(\sum_{k=1}^m \frac{r_k^4}{w_k^3} - 1 \right) \\ &= \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{2\rho_1 \rho_2^2} \left(\sum_{k=1}^m \frac{1}{w_k (k + \theta)^{2s} M_{m,\theta,s}^2} - 1 \right) \\ & - \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{6\rho_1 \rho_2^3} \left(\sum_{k=1}^m \frac{1}{w_k^2 (k + \theta)^{3s} M_{m,\theta,s}^3} - 1 \right) \\ & + \frac{1}{12\rho_2^3} \left(\sum_{k=1}^m \frac{1}{w_k^3 (k + \theta)^{4s} M_{m,\theta,s}^4} - 1 \right). \end{aligned} \quad (42)$$

Now using (41) and (42) in (36), we get (40). \square

Corollary 10 Let $0 < \rho_1 < \rho_2$, $\theta_1, \theta_2 \geq 0$, $s_1, s_2 > 0$ with $(k + \theta_2)^{s_2} M_{m,\theta_2,s_2} / (k + \theta_1)^{s_1} M_{m,\theta_1,s_1} \in [\rho_1, \rho_2]$ for $k = 1, \dots, m$, then

$$\begin{aligned} & -Z(M, \theta_1, s_1) + \sum_{k=1}^m \frac{\log(k + \theta_2)^{s_2} M_{m,\theta_2,s_2}}{(k + \theta_1)^{s_1} M_{m,\theta_1,s_1}} \\ & \leq \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{2\rho_1 \rho_2^2} \left(\sum_{k=1}^m \frac{(k + \theta_2)^{s_2} M_{m,\theta_2,s_2}}{(k + \theta_1)^{2s_1} M_{m,\theta_1,s_1}^2} - 1 \right) \\ & - \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{6\rho_1 \rho_2^3} \left(\sum_{k=1}^m \frac{(k + \theta_2)^{2s_2} M_{m,\theta_2,s_2}^2}{(k + \theta_1)^{3s_1} M_{m,\theta_1,s_1}^3} - 1 \right) \\ & + \frac{1}{12\rho_2^3} \left(\sum_{k=1}^m \frac{(k + \theta_2)^{3s_2} M_{m,\theta_2,s_2}^3}{(k + \theta_1)^{4s_1} M_{m,\theta_1,s_1}^4} - 1 \right). \end{aligned} \quad (43)$$

Proof: For $r_k = 1/(k + \theta_1)^{s_1} M_{m,\theta_1,s_1}$, $w_k = 1/(k + \theta_2)^{s_2} M_{m,\theta_2,s_2}$, $k = 1, \dots, m$, we have

$$\begin{aligned} & \sum_{k=1}^m r_k \log \frac{r_k}{w_k} = \sum_{k=1}^m \frac{1}{(k + \theta_1)^{s_1} M_{m,\theta_1,s_1}} \\ & \times \left(\log(k + \theta_2)^{s_2} M_{m,\theta_2,s_2} - \log(k + \theta_1)^{s_1} M_{m,\theta_1,s_1} \right) \\ & = -Z(M, \theta_1, s_1) + \sum_{k=1}^m \frac{\log(k + \theta_2)^{s_2} M_{m,\theta_2,s_2}}{(k + \theta_1)^{s_1} M_{m,\theta_1,s_1}}. \end{aligned} \quad (44)$$

Also,

$$\begin{aligned} & \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{2\rho_1 \rho_2^2} \left(\sum_{k=1}^m \frac{r_k^2}{w_k} - 1 \right) \\ & - \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{6\rho_1 \rho_2^3} \left(\sum_{k=1}^m \frac{r_k^3}{w_k^2} - 1 \right) + \frac{1}{12\rho_2^3} \left(\sum_{k=1}^m \frac{r_k^4}{w_k^3} - 1 \right) \\ &= \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{2\rho_1 \rho_2^2} \left(\sum_{k=1}^m \frac{(k + \theta_2)^{s_2} M_{m,\theta_2,s_2}}{(k + \theta_1)^{2s_1} M_{m,\theta_1,s_1}^2} - 1 \right) \\ & - \frac{\rho_1^2 + \rho_1 \rho_2 + \rho_2^2}{6\rho_1 \rho_2^3} \left(\sum_{k=1}^m \frac{(k + \theta_2)^{2s_2} M_{m,\theta_2,s_2}^2}{(k + \theta_1)^{3s_1} M_{m,\theta_1,s_1}^3} - 1 \right) \\ & + \frac{1}{12\rho_2^3} \left(\sum_{k=1}^m \frac{(k + \theta_2)^{3s_2} M_{m,\theta_2,s_2}^3}{(k + \theta_1)^{4s_1} M_{m,\theta_1,s_1}^4} - 1 \right). \end{aligned} \quad (45)$$

Now utilizing (44) and (45) in (36), we get (43). \square

CONCLUSION

In this paper, we have utilized strongly convex functions and improved an earlier bound for Jensen’s gap in its discrete form which is presented in [18]. Also, we have presented the improvement of an earlier bound for Jensen’s gap of its integral form which is introduced in [17]. As consequences of these improved results, we have achieved some improvements of the earlier results presented in [17, 18] around various divergences, the Hermite-Hadamard and Hölder inequalities. We have demonstrated these improvements through numerical experiments. Finally, we have provided some new results around the Zipf-Mandelbrot entropy. The idea of the paper may strengthen some other existing results based on the notion of classical convexity.

REFERENCES

1. Pečarić J, Proschan F, Tong YL (1992) *Convex Functions, Partial Orderings and Statistical Applications*, Academic Press, New York.
2. Dragomir SS, Pearce CEM, Pečarić J (2003) Interpolations of Jensen’s inequality. *Tamkang J Math* **34**, 175–188.
3. Kim JH (2016) Further improvement of Jensen inequality and application to stability of time-delayed systems. *Automatica* **64**, 121–125.
4. Zhang D, Shen YP, Zhou SQ, Dong XW, Yu L (2021) Distributed secure platoon control of connected vehicles subject to Dos attack: theory and applications. *IEEE Trans Syst Man Cybern Syst* **51**, 7269–7278.
5. Dragomir SS (2013) Some reverses of the Jensen inequality with applications. *Bull Aust Math Soc* **87**, 177–194.
6. Khalid S, Pečarić D, Pečarić J (2019) On Shannon and Zipf-Mandelbrot entropies and related results. *J Inequal Appl* **2019**, 99.
7. Kluz PA (2020) On Jensen-Rényi and Jeffreys-Rényi type f -divergences induced by convex functions. *Phys A Stat Mech Appl* **548**, 122527.

8. Mehmood N, Butt SI, Pečarić D, Pečarić J (2020) Generalizations of cyclic refinements of Jensen's inequality by Lidstone's polynomial with applications in information theory. *J Math Inequal* **14**, 249–271.
9. Pečarić J, Perić J (2020) Refinements of the integral form of Jensen's and the Lah-Ribarić inequalities and applications for Csizsár divergence. *J Inequal Appl* **2020**, 108.
10. Bakula MK, Pečarić J, Ribičić M (2006) Companion inequalities to Jensen's inequality for m -convex and (α, m) -convex functions. *J Inequal Pure Appl Math* **7**, 194.
11. Dragomir SS, Fitzpatrick S (2000) The Jensen inequality for s -breckner convex functions in linear spaces. *Demonstratio Math* **33**, 43–49.
12. Mihai MV, Niculescu CP (2016) A simple proof of the Jensen-type inequality of Fink and Jodeit. *Mediterr J Math* **13**, 119–126.
13. Moslehian MS, Kian M (2012) Jensen type inequalities for Q -class functions. *Bull Aust Math Soc* **85**, 128–142.
14. Moradi HR, Omidvar ME, Adil Khan M, Nikodem K (2018) Around Jensen's inequality for strongly convex functions. *Aequat Math* **92**, 25–37.
15. Noor MA (2004) Some developments in general variational inequalities. *Appl Math Comput* **152**, 199–277.
16. Youness EA (1999) E-convex sets, E-convex functions, and E-convex programming. *J Optim Theory Appl* **102**, 439–450.
17. Adil Khan M, Khan S, Chu YM (2020) A new bound for the Jensen gap with applications in information theory. *IEEE Access* **8**, 98001–98008.
18. Khan S, Adil Khan M, Butt SI, Chu YM (2020) A new bound for the Jensen gap pertaining twice differentiable functions with applications. *Adv Differ Equ* **2020**, 333.
19. Polyak BT (1966) Existence theorems and convergence of minimizing sequences in extremum problems with restrictions. *Soviet Math Dokl* **7**, 72–75.
20. Merentes N, Nikodem K (2010) Remarks on strongly convex functions. *Aequat Math* **80**, 193–199.
21. Pečarić D, Pečarić J, Rodić M (2018) About the sharpness of the Jensen inequality. *J Inequal Appl* **2018**, 337.
22. Shannon CE (1951) Prediction and entropy of printed English. *Bell Sys Tech J* **30**, 50–64.
23. Zipf GK (1949) *Human Behavior and the Principle of Least Effort*, Addison-Wesley Press.
24. Mandelbrot B (1965) Information theory and psycholinguistics: a theory of word frequencies. In: Lazarsfeld PF, Henry NW (eds) *Readings in Mathematical Social Science*, Science Research Associates, Chicago, pp 350–368.