Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH

Open Access

Generalized Jensen and Jensen–Mercer inequalities for strongly convex functions with applications

Slavica Ivelić Bradanović[1] and Neda Lovričević[1*]

*Correspondence:
neda.lovricevic@gradst.hr
[1]Faculty of Civil Engineering, Architecture and Geodesy, University of Split, Matice hrvatske 15, 21000 Split, Croatia

Abstract

Strongly convex functions as a subclass of convex functions, still equipped with stronger properties, are employed through several generalizations and improvements of the Jensen inequality and the Jensen–Mercer inequality. This paper additionally provides applications of obtained main results in the form of new estimates for so-called strong f-divergences: the concept of the Csiszár f-divergence for strongly convex functions f, together with particular cases (Kullback–Leibler divergence, χ^2-divergence, Hellinger divergence, Bhattacharya distance, Jeffreys distance, and Jensen–Shannon divergence.) Furthermore, new estimates for the Shannon entropy are obtained, and new Chebyshev-type inequalities are derived.

Mathematics Subject Classification: 26D15; 26D20; 94A15; 94A17

Keywords: Convex and strongly convex functions; Jensen inequality; Jensen–Mercer inequality; Chebyshev functional; Strong f-divergences

1 Introduction

One of the extended approaches to convexity developed in the last century includes strongly convex functions as a subclass of convex functions (see [20] and for more recent contributions, [10, 11, 18]).

Let us recall that a function $f: [a,b] \subseteq \mathbb{R} \to \mathbb{R}$ is strongly convex with modulus $c > 0$ if

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) - c\lambda(1-\lambda)(x-y)^2 \tag{1.1}$$

for all $x, y \in [a,b]$ and $\lambda \in [0,1]$.

A function f that satisfies (1.1) with $c = 0$, i.e.,

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y), \tag{1.2}$$

is convex in the usual sense. Obviously, strong convexity implies convexity, but the reverse implication is not true in general. For example, a linear function is convex but is not strongly convex.

Comparing with convex functions, the strongly convex ones possess stronger versions of the analogous properties. One of their useful characterizations is given in the following lemma (see [23, p. 268], [11, 20], and the references therein).

Lemma 1 *A function $f: [a,b] \to \mathbb{R}$ is strongly convex with modulus $c > 0$ iff the function $g: [a,b] \to \mathbb{R}$ defined by $g(x) = f(x) - cx^2$ is convex.*

We further use the well-known theorem proved by Stolz [19, p. 25].

Theorem 1 (Stolz) *Let $f: [a,b] \to \mathbb{R}$ be a convex function. Then f is continuous on (a,b) and has finite left and right derivatives at each point of (a,b). Both f'_- and f'_+ are nondecreasing on (a,b). Moreover, for all $x,y \in (a,b)$, $x < y$, we have*

$$f'_-(x) \le f'_+(x) \le f'_-(y) \le f'_+(y).$$

Strongly convex functions are accompanied by the corresponding Jensen inequality, which was proved in [20].

Theorem 2 *Let a function $f: (a,b) \to \mathbb{R}$ be strongly convex with modulus $c > 0$. Suppose $\boldsymbol{x} = (x_1, \dots, x_n) \in (a,b)^n$ and $\boldsymbol{a} = (a_1, \dots, a_n)$ is a nonnegative n-tuple such that $A_n = \sum_{i=1}^n a_i > 0$ with $\bar{x} = \frac{1}{A_n}\sum_{i=1}^n a_i x_i$. Then*

$$f(\bar{x}) \le \frac{1}{A_n}\sum_{i=1}^n a_i f(x_i) - \frac{c}{A_n}\sum_{i=1}^n a_i (x_i - \bar{x})^2. \tag{1.3}$$

It is easily seen that for $c = 0$, inequality (1.3) becomes the Jensen inequality for convex functions:

$$f(\bar{x}) \le \frac{1}{A_n}\sum_{i=1}^n a_i f(x_i). \tag{1.4}$$

Inequality (1.3) provides a better upper bound for $f(\bar{x})$ because of the nonnegativity of the term $\frac{c}{A_n}\sum_{i=1}^n a_i (x_i - \bar{x})^2$. Thus (1.3) is an improvement of (1.4) and is considered as its stronger variant.

Another Jensen-type inequality was established by Mercer [17]. Given a convex function $f: (a,b) \to \mathbb{R}$ with $m, M \in (a,b)$, $m < M$, for $\boldsymbol{x} = (x_1, \dots, x_n) \in [m,M]^n$ and a nonnegative n-tuple $\boldsymbol{a} = (a_1, \dots, a_n)$ such that $A_n = \sum_{i=1}^n a_i > 0$ with $\bar{x} = \frac{1}{A_n}\sum_{i=1}^n a_i x_i$, the Jensen–Mercer inequality states that

$$f(m + M - \bar{x}) \le f(m) + f(M) - \frac{1}{A_n}\sum_{i=1}^n a_i f(x_i). \tag{1.5}$$

Numerous improvements and generalizations of (1.5) have been obtained since. Here we accentuate two such results. In [15] the authors proved that for a convex function $f: (a,b) \to \mathbb{R}$, $\boldsymbol{x} = (x_1, \dots, x_n) \in [m,M]^n$, where $m, M \in (a,b)$, $m < M$, and a nonnegative n-tuple $\boldsymbol{a} = (a_1, \dots, a_n)$ such that $A_n = \sum_{i=1}^n a_i > 0$, we have the inequalities

$$f(c) + f'(c)\left(m + M - c - \frac{1}{A_n}\sum_{i=1}^n a_i x_i\right) \tag{1.6}$$

$$\le f(m)+f(M)-\frac{1}{A_n}\sum_{i=1}^{n}a_if(x_i)$$

$$\le f(d)+f'(m)(m-d)+f'(M)(M-d)-\frac{1}{A_n}\sum_{i=1}^{n}a_if'(x_i)(x_i-d)$$

for all $c,d\in[m,M]$.

Furthermore, the following variant of the Jensen–Mercer inequality was proved in [18] for strongly convex functions.

Theorem 3 *Let $f:(a,b)\to\mathbb{R}$ be a strongly convex function, and let $m,M\in(a,b)$, $m<M$. Let $\boldsymbol{x}=(x_1,\dots,x_n)\in[m,M]^n$, and let $\boldsymbol{a}=(a_1,\dots,a_n)$ be a nonnegative n-tuple such that $\sum_{i=1}^{n}a_i=1$ with $\bar{x}=\sum_{i=1}^{n}a_ix_i$. Let $\lambda_i\in[0,1]$, $i\in\{1,\dots,n\}$. Then*

$$f(m+M-\bar{x})\le f(m)+f(M)-\sum_{i=1}^{n}a_if(x_i) \\ -c\left[2(M-m)^2\sum_{i=1}^{n}a_i\lambda_i(1-\lambda_i)+\sum_{i=1}^{n}a_i(\bar{x}-x_i)^2\right]. \tag{1.7}$$

For some recent results on the Jensen–Mercer inequality, see [1–3, 9, 12–14, 16, 24].

With the aim of new improvements and elaborating the existing results, the paper is divided into five sections. In Section 1, we recall a few results needed further: some on strongly convex functions and some well-known ones, concerning convex functions. Sections 2 and 3 deal with the Jensen and Jensen–Mercer inequalities, both generalized by means of strongly convex functions. In Sect. 4, we discuss applications to Csiszár strong f-divergences introduced in [10], for which we provide new estimates and their particular types in the same manner. We also derive new estimates for the Shannon entropy. Section 5 deals with new Chebyshev-type inequalities.

2 The Jensen-type inequalities

We start this section with important properties of strongly convex functions, which are direct consequences of the characterizations given in Lemma 1 and Theorem 1.

Lemma 2 *Let $f:[a,b]\to\mathbb{R}$ be a strongly convex function with modulus $c>0$. Then it is continuous on (a,b) and has finite left and right derivatives at each point of (a,b). Both f'_- and f'_+ are nondecreasing on (a,b). Moreover, for all $x,y\in(a,b)$, $x<y$, we have*

$$f'_-(x)-2cx\le f'_+(x)-2cx\le f'_-(y)-2cy\le f'_+(y)-2cy. \tag{2.1}$$

If f is differentiable, then f' is strongly increasing on (a,b), i.e., for all $x,y\in(a,b)$, $x<y$,

$$f'(x)+2c(y-x)\le f'(y). \tag{2.2}$$

Proof Let id denote the identity function, i.e., $id(t)=t$ for all $t\in[a,b]$. Since f is strongly convex with modulus $c>0$, the function $g=f-c\cdot id^2$ is convex. Now, as an easy consequence of Theorem 1 applied to the convex function $g=f-c\cdot id^2$, we get the first part of the statement.

If f is differentiable, then $f'(x) = f'_-(x) = f'_+(x)$ and $f'(y) = f'_-(y) = f_+(y)$, and (2.1) implies (2.2). □

Bearing in mind the statement of the previous lemma, for a strongly convex function $f: [a,b] \to \mathbb{R}$, by $f'(x)$, $x \in (a,b)$, we mean that $f'(x)$ is any element from the interval $[f'_-(x), f'_+(x)]$. If f is differentiable, then $f'(x) = f'_-(x) = f'_+(x)$.

Furthermore, for a strongly convex function $f: [a,b] \to \mathbb{R}$ with modulus $c > 0$, we have

$$f(x) \geq f(y) + f'(y)(x-y) + c(x-y)^2 \tag{2.3}$$

for all $x, y \in (a,b)$. This inequality is as an easy consequence of the characterization of convex functions via support lines (see [21, Theorem 1.6]) applied to the convex function $g = f - c \cdot id^2$.

A generalization and an improvement of Jensen's inequality (1.3) for strongly convex functions is included in the following theorem.

Theorem 4 *Let $f: (a,b) \to \mathbb{R}$ be a strongly convex function with modulus $c > 0$. Suppose $\boldsymbol{x} = (x_1, \dots, x_n) \in (a,b)^n$ and $\boldsymbol{a} = (a_1, \dots, a_n)$ is a nonnegative n-tuple with $A_n = \sum_{i=1}^n a_i > 0$. Let $\bar{x} = \frac{1}{A_n} \sum_{i=1}^n a_i x_i$ and $\hat{x}_i = (1-\lambda_i)\bar{x} + \lambda_i x_i$, $\lambda_i \in [0,1]$, $i \in \{1, \dots, n\}$. Then*

$$\begin{aligned}
0 &\leq \left| \frac{1}{A_n} \sum_{i=1}^n a_i \left| f(x_i) - f(\hat{x}_i) - c(1-\lambda_i)^2(\bar{x}-x_i)^2 \right| - \frac{1}{A_n} \sum_{i=1}^n a_i (1-\lambda_i) \left| f'(\hat{x}_i) \right| |\bar{x}-x_i| \right| \\
&\leq \frac{1}{A_n} \sum_{i=1}^n a_i f(x_i) - \frac{1}{A_n} \sum_{i=1}^n a_i f(\hat{x}_i) \\
&\quad - \frac{1}{A_n} \sum_{i=1}^n a_i (1-\lambda_i) f'(\hat{x}_i)(\bar{x}-x_i) - \frac{c}{A_n} \sum_{i=1}^n a_i (1-\lambda_i)^2 (\bar{x}-x_i)^2.
\end{aligned} \tag{2.4}$$

Proof Applying the triangle inequality $||u| - |v|| \leq |u - v|$ to (2.3), we get

$$\begin{aligned}
&\left| \left| f(x) - f(y) - c(x-y)^2 \right| - \left| f'(y) \right| \left| (x-y) \right| \right| \\
&\leq \left| f(x) - f(y) - c(x-y)^2 - f'(y)(x-y) \right| \\
&= f(x) - f(y) - c(x-y)^2 - f'(y)(x-y).
\end{aligned} \tag{2.5}$$

Setting $y = \hat{x}_i$ and $x = x_i$, $i \in \{1, \dots, n\}$, from (2.5) we have

$$\begin{aligned}
&\left| \left| f(x_i) - f(\hat{x}_i) - c(1-\lambda_i)^2(\bar{x}-x_i)^2 \right| - (1-\lambda_i) \left| f'(\hat{x}_i) \right| |\bar{x}-x_i| \right| \\
&\leq \left| f(x_i) - f(\hat{x}_i) - (1-\lambda_i) f'(\hat{x}_i)(\bar{x}-x_i) - c(1-\lambda_i)^2(\bar{x}-x_i)^2 \right| \\
&= f(x_i) - f(\hat{x}_i) - (1-\lambda_i) f'(\hat{x}_i)(\bar{x}-x_i) - c(1-\lambda_i)^2(\bar{x}-x_i)^2.
\end{aligned}$$

Now multiplying by a_i, summing over i, $i = 1, \dots, n$, and then dividing by $A_n = \sum_{i=1}^n a_i > 0$, we get

$$\frac{1}{A_n} \sum_{i=1}^n a_i \left| \left| f(x_i) - f(\hat{x}_i) - c(1-\lambda_i)^2(\bar{x}-x_i)^2 \right| - (1-\lambda_i) \left| f'(\hat{x}_i) \right| |\bar{x}-x_i| \right|$$

$$
\begin{aligned}
&\leq \frac{1}{A_n}\sum_{i=1}^{n} a_i \left| f(x_i) - f(\hat{x}_i) - (1-\lambda_i) f'(\hat{x}_i)(\bar{x}-x_i) - c(1-\lambda_i)^2(\bar{x}-x_i)^2 \right| \\
&= \frac{1}{A_n}\sum_{i=1}^{n} a_i f(x_i) - \frac{1}{A_n}\sum_{i=1}^{n} a_i f(\hat{x}_i) \\
&- \frac{1}{A_n}\sum_{i=1}^{n} a_i (1-\lambda_i) f'(\hat{x}_i)(\bar{x}-x_i) - \frac{c}{A_n}\sum_{i=1}^{n} a_i (1-\lambda_i)^2(\bar{x}-x_i)^2. \qquad (2.6)
\end{aligned}
$$

By the triangle inequality ($\left|\sum_{i=1}^{n} a_i z_i\right| \leq \sum_{i=1}^{n} a_i |z_i|$), we also have

$$
\begin{aligned}
&\left| \frac{1}{A_n}\sum_{i=1}^{n} a_i \left| f(x_i) - f(\hat{x}_i) - c(1-\lambda_i)^2(\bar{x}-x_i)^2 \right| - \frac{1}{A_n}\sum_{i=1}^{n} a_i (1-\lambda_i) \left| f'(\hat{x}_i) \right| |\bar{x}-x_i| \right| \\
&\leq \frac{1}{A_n}\sum_{i=1}^{n} a_i \left| \left| f(x_i) - f(\hat{x}_i) - c(1-\lambda_i)^2(\bar{x}-x_i)^2 \right| - (1-\lambda_i) \left| f'(\hat{x}_i) \right| |\bar{x}-x_i| \right|. \qquad (2.7)
\end{aligned}
$$

Now combining (2.6) and (2.7), we get (2.4). □

The following corollary is a direct consequence of Theorem 4.

Corollary 1 *Let $f: (a,b) \to \mathbb{R}$ be a strongly convex function with modulus $c > 0$. Suppose $\boldsymbol{x} = (x_1, \dots, x_n) \in (a,b)^n$ and $\boldsymbol{a} = (a_1, \dots, a_n)$ is a nonnegative n-tuple with $A_n = \sum_{i=1}^{n} a_i > 0$ and $\bar{x} = \frac{1}{A_n}\sum_{i=1}^{n} a_i x_i$. Then*

$$
\begin{aligned}
0 &\leq \left| \frac{1}{A_n}\sum_{i=1}^{n} a_i \left| f(x_i) - f(\bar{x}) - c(x_i-\bar{x})^2 \right| - \left| f'(\bar{x}) \right| \cdot \frac{1}{A_n}\sum_{i=1}^{n} a_i |x_i-\bar{x}| \right| \\
&\leq \frac{1}{A_n}\sum_{i=1}^{n} a_i f(x_i) - f(\bar{x}) - \frac{c}{A_n}\sum_{i=1}^{n} a_i (x_i-\bar{x})^2. \qquad (2.8)
\end{aligned}
$$

Proof Setting $\lambda_i = 0$, $i = 1, \dots, n$, from (2.4) we get

$$
\begin{aligned}
0 &\leq \left| \frac{1}{A_n}\sum_{i=1}^{n} a_i \left| f(x_i) - f(\bar{x}) - c(\bar{x}-x_i)^2 \right| - \frac{1}{A_n}\sum_{i=1}^{n} a_i \left| f'(\bar{x}) \right| |\bar{x}-x_i| \right| \qquad (2.9) \\
&\leq \frac{1}{A_n}\sum_{i=1}^{n} a_i f(x_i) - \frac{1}{A_n}\sum_{i=1}^{n} a_i f(\bar{x}) \\
&- \frac{1}{A_n}\sum_{i=1}^{n} a_i f'(\bar{x})(\bar{x}-x_i) - \frac{c}{A_n}\sum_{i=1}^{n} a_i (\bar{x}-x_i)^2.
\end{aligned}
$$

Note that

$$
\frac{1}{A_n}\sum_{i=1}^{n} a_i f'(\bar{x})(x_i-\bar{x}) = f'(\bar{x})\frac{1}{A_n}\sum_{i=1}^{n} a_i (x_i-\bar{x}) = 0. \qquad (2.10)
$$

Now combining (2.9) and (2.10), we get (2.8). □

Finally, in a similar manner, we get an inequality, which counterparts the Jensen inequality (1.3).

Theorem 5 *Let $f\colon (a,b) \to \mathbb{R}$ be a strongly convex function with modulus $c > 0$. Suppose $\boldsymbol{x} = (x_1,\dots,x_n) \in (a,b)^n$ and $\boldsymbol{a} = (a_1,\dots,a_n)$ is a nonnegative n-tuple with $A_n = \sum_{i=1}^n a_i > 0$ and $\bar{x} = \frac{1}{A_n}\sum_{i=1}^n a_i x_i$. Let $\lambda_i \in [0,1]$, $i \in \{1,\dots,n\}$. Then*

$$\begin{aligned} &\frac{1}{A_n}\sum_{i=1}^n a_i f(x_i) - \frac{1}{A_n}\sum_{i=1}^n a_i f\big((1-\lambda_i)\bar{x} + \lambda_i x_i\big) \\ &\quad\le \frac{1}{A_n}\sum_{i=1}^n a_i(1-\lambda_i) f'(x_i)(x_i - \bar{x}) - \frac{c}{A_n}\sum_{i=1}^n a_i (1-\lambda_i)^2 (\bar{x} - x_i)^2. \end{aligned} \tag{2.11}$$

Proof With (2.3) slightly modified, for $x = (1-\lambda_i)\bar{x} + \lambda_i x_i$ and $y = y_i$, $i \in \{1,\dots,n\}$, we have

$$f\big((1-\lambda_i)\bar{x} + \lambda_i x_i\big) - f(x_i) \ge f'(x_i)(1-\lambda_i)(\bar{x} - x_i) + c(1-\lambda_i)^2(\bar{x} - x_i)^2.$$

Now multiplying by a_i, summing over i, $i = 1,\dots,n$, and then dividing by $A_n > 0$, we get

$$\begin{aligned} &\frac{1}{A_n}\sum_{i=1}^n a_i f\big((1-\lambda_i)\bar{x} + \lambda_i x_i\big) - \frac{1}{A_n}\sum_{i=1}^n a_i f(x_i) \\ &\quad\ge \frac{1}{A_n}\sum_{i=1}^n a_i(1-\lambda_i) f'(x_i)(\bar{x} - x_i) + \frac{c}{A_n}\sum_{i=1}^n a_i (1-\lambda_i)^2 (\bar{x} - x_i)^2, \end{aligned}$$

which is equivalent to (2.11). □

Again, a direct consequence of Theorem 5 follows by setting $\lambda_i = 0$ for $i = 1,\dots,n$.

Corollary 2 *Let $f\colon (a,b) \to \mathbb{R}$ be a strongly convex function with modulus $c > 0$. Suppose $\boldsymbol{x} = (x_1,\dots,x_n) \in (a,b)^n$ and $\boldsymbol{a} = (a_1,\dots,a_n)$ is a nonnegative n-tuple with $A_n = \sum_{i=1}^n a_i > 0$ and $\bar{x} = \frac{1}{A_n}\sum_{i=1}^n a_i x_i$. Then*

$$\begin{aligned} 0 &\le \frac{1}{A_n}\sum_{i=1}^n a_i f(x_i) - f(\bar{x}) \\ &\le \frac{1}{A_n}\sum_{i=1}^n a_i f'(x_i) x_i - \frac{1}{A_n^2}\sum_{i=1}^n a_i x_i \sum_{i=1}^n a_i f'(x_i) - \frac{c}{A_n}\sum_{i=1}^n a_i(\bar{x} - x_i)^2. \end{aligned} \tag{2.12}$$

Remark 1 Our results generalize and improve the main results obtained in [7, 8], which were related to convex functions.

3 The Jensen–Mercer-type inequalities

We embark on further investigation of the Jensen–Mercer inequality (1.5). Along the way, we generalize and improve results (1.6) from [15] and (1.7) from [18].

Theorem 6 *Let a function $f\colon (a,b) \to \mathbb{R}$ be strongly convex with modulus $c > 0$, and let $m, M \in (a,b)$, $m < M$, and $\lambda_i \in [0,1]$, $i \in \{1,\dots,n\}$. Suppose $\boldsymbol{x} = (x_1,\dots,x_n) \in [m,M]^n$ and $\boldsymbol{a} = (a_1,\dots,a_n)$ is a nonnegative n-tuple with $A_n = \sum_{i=1}^n a_i > 0$ and $\bar{x} = \frac{1}{A_n}\sum_{i=1}^n a_i x_i$ Then*

$$f(d) + f'(d)\,(m + M - d - \bar{x}) + \frac{c}{A_n}\sum_{i=1}^n a_i (m + M - d - x_i)^2$$

$$
\begin{aligned}
&+\frac{2c(M-m)^2}{A_n}\sum_{i=1}^{n}a_i\lambda_i(1-\lambda_i) \qquad (3.1)\\
&\leq f(m)+f(M)-\frac{1}{A_n}\sum_{i=1}^{n}a_if(x_i)\\
&\leq f(e)+f'(m)(m-e)+f'(M)(M-e)-\frac{1}{A_n}\sum_{i=1}^{n}a_if'(x_i)(x_i-e)\\
&-\frac{c}{A_n}\sum_{i=1}^{n}a_i(M-x_i)(3x_i-2e-M)+c(m-e)^2
\end{aligned}
$$

for all $d,e\in[m,M]$.

Proof Let $\lambda_i\in[0,1]$, $x_i\in[m,M]$, and $y_i=m+M-x_i$, $i\in\{1,\dots,n\}$. Then we can write as convex combinations:

$$
\begin{aligned}
&x_i=\lambda_im+(1-\lambda_i)M,\\
&y_i=(1-\lambda_i)m+\lambda_iM,\quad i\in\{1,\dots,n\}.
\end{aligned}
$$

Applying (1.1) twice, we have

$$
\begin{aligned}
f(m+M-x_i)&=f((1-\lambda_i)m+\lambda_iM)\\
&\leq(1-\lambda_i)f(m)+\lambda_if(M)-c\lambda_i(1-\lambda_i)(M-m)^2\\
&=f(m)+f(M)-\lambda_if(m)+\lambda_if(M)-f(M)-c\lambda_i(1-\lambda_i)(M-m)^2\\
&=f(m)+f(M)-\left[\lambda_if(m)+(1-\lambda_i)f(M)\right]-c\lambda_i(1-\lambda_i)(M-m)^2\\
&\leq f(m)+f(M)-f(\lambda_im+(1-\lambda_i)M)-2c\lambda_i(1-\lambda_i)(M-m)^2\\
&=f(m)+f(M)-f(x_i)-2c\lambda_i(1-\lambda_i)(M-m)^2.
\end{aligned}
$$

Further, applying (2.3), we get

$$
f(d)+f'(d)(m+M-x_i-d)+c(m+M-x_i-d)^2\leq f(m+M-x_i),
$$

which, combined with the previous inequality, implies

$$
\begin{aligned}
&f(d)+f'(d)(m+M-x_i-d)+c(m+M-x_i-d)^2 \qquad (3.2)\\
&\leq f(m+M-x_i)\\
&\leq f(m)+f(M)-f(x_i)-2c\lambda_i(1-\lambda_i)(M-m)^2.
\end{aligned}
$$

Furthermore, for $x_i,e\in[m,M]$, $i\in\{1,\dots,n\}$, by (2.3) we have

$$
\begin{aligned}
&f(m)-f(e)\leq f'(m)(m-e)+c(m-e)^2, \qquad (3.3)\\
&f(M)-f(x_i)\leq f'(M)(M-x_i)+c(M-x_i)^2.
\end{aligned}
$$

Using (3.3), we have

$$\begin{aligned} & f(m)+f(M)-f(x_i)-2c\lambda_i(1-\lambda_i)(M-m)^2 \\ & = f(e)+f(m)-f(e)+f(M)-f(x_i)-2c\lambda_i(1-\lambda_i)(M-m)^2 \\ & \leq f(e)+f'(m)(m-e)+f'(M)(M-x_i) \\ & + c(m-e)^2+c(M-x_i)^2-2c\lambda_i(1-\lambda_i)(M-m)^2 \\ & = f(e)+f'(m)(m-e)+f'(M)(M-e)-f'(M)(x_i-e) \\ & + c(m-e)^2+c(M-x_i)^2-2c\lambda_i(1-\lambda_i)(M-m)^2. \end{aligned} \tag{3.4}$$

Since f' is strongly increasing and $x_i \leq M$, by (2.2) we have $-f'(M) \leq -f'(x_i)-2c(M-x_i)$, i.e.,

$$\begin{aligned} & f(e)+f'(m)(m-e)+f'(M)(M-e)-f'(M)(x_i-e) \\ & + c(m-e)^2+c(M-x_i)^2-2c\lambda_i(1-\lambda_i)(M-m)^2 \\ & \leq f(e)+f'(m)(m-e)+f'(M)(M-e)-\left[f'(x_i)+2c(M-x_i)\right](x_i-e) \\ & + c(m-e)^2+c(M-x_i)^2-2c\lambda_i(1-\lambda_i)(M-m)^2. \end{aligned} \tag{3.5}$$

Combining (3.4) and (3.5), we get

$$\begin{aligned} & f(m)+f(M)-f(x_i)-2c\lambda_i(1-\lambda_i)(M-m)^2 \\ & \leq f(e)+f'(m)(m-e)+f'(M)(M-e)-\left[f'(x_i)+2c(M-x_i)\right](x_i-e) \\ & + c(m-e)^2+c(M-x_i)^2-2c\lambda_i(1-\lambda_i)(M-m)^2 \\ & = f(e)+f'(m)(m-e)+f'(M)(M-e) \\ & - f'(x_i)(x_i-e)-2c(M-x_i)(x_i-e) \\ & + c(m-e)^2+c(M-x_i)^2-2c\lambda_i(1-\lambda_i)(M-m)^2. \end{aligned} \tag{3.6}$$

Finally, from (3.2) and (3.6) we have

$$\begin{aligned} & f(d)+f'(d)(m+M-x_i-d)+c(m+M-x_i-d)^2 \\ & \leq f(m)+f(M)-f(x_i)-2c\lambda_i(1-\lambda_i)(M-m)^2 \\ & \leq f(e)+f'(m)(m-e)+f'(M)(M-e) \\ & - f'(x_i)(x_i-e)-2c(M-x_i)(x_i-e) \\ & + c(m-e)^2+c(M-x_i)^2-2c\lambda_i(1-\lambda_i)(M-m)^2. \end{aligned}$$

Multiplying it by a_i, summing over $i, i=1,\dots,n$, and then dividing by $A_n > 0$, we get (3.1). □

Remark 2 In particular, if we set $A_n = 1$ and $d = m+M-\bar{x}$, then the first inequality in (3.1) becomes (1.7) from [18], which makes it a generalization. Furthermore, our result (3.1) improves (1.6) from [15].

As an easy consequence of the previous theorem, we get the following inequality of the Jensen–Mercer type.

Corollary 3 *Let the assumptions of Theorem 6 hold. Then*

$$f(m+M-\bar{x})+\frac{c}{A_n}\sum_{i=1}^{n}a_i(\bar{x}-x_i)^2+\frac{2c(M-m)^2}{A_n}\sum_{i=1}^{n}a_i\lambda_i(1-\lambda_i) \tag{3.7}$$
$$\leq f(m)+f(M)-\frac{1}{A_n}\sum_{i=1}^{n}a_if(x_i)$$
$$\leq f(\bar{x})+f'(m)(m-\bar{x})+f'(M)(M-\bar{x})-\frac{1}{A_n}\sum_{i=1}^{n}a_if'(x_i)(x_i-\bar{x})$$
$$-\frac{c}{A_n}\sum_{i=1}^{n}a_i(M-x_i)(3x_i-2\bar{x}-M)+c(m-\bar{x})^2.$$

Proof Choosing $e=\bar{x}=\frac{1}{A_n}\sum_{i=1}^{n}a_ix_i$ and $d=m+M-\bar{x}$, from (3.1) we get (3.7). □

4 Applications to strong f-divergences and the Shannon entropy

Let $\mathcal{P}_n=\left\{\mathbf{p}=(p_1,\dots,p_n)\colon p_1,\dots,p_n>0,\sum_{i=1}^{n}p_i=1\right\}$ be the set of all complete finite discrete probability distributions. The restriction to positive distributions is only for convenience. If we take $p_i=0$ for some $i\in\{1,\dots,n\}$, then in the following results, we need to interpret undefined expressions as $f(0)=\lim_{t\to0+}f(t)$, $0f\left(\frac{0}{0}\right)=0$, and $0f\left(\frac{e}{0}\right)=\lim_{\varepsilon\to0+}f\left(\frac{e}{\varepsilon}\right)=e\lim_{t\to\infty}\frac{f(t)}{t}$, $e>0$.

I. Csiszár [5] introduced an important class of statistical divergences by means of convex functions.

Definition 1 Let $f\colon(0,\infty)\to\mathbb{R}$ be a convex function, and let $\mathbf{p},\mathbf{q}\in\mathcal{P}_n$. The Csiszár f-divergence is defined as

$$D_f(\mathbf{q},\mathbf{p})=\sum_{i=1}^{n}p_if\left(\frac{q_i}{p_i}\right). \tag{4.1}$$

It has deep and fruitful applications in various branches of science (see, e.g., [4, 22] with references therein) and is involved in the following Csiszár–Körner inequality (see [6]).

Theorem 7 *Let $\mathbf{p},\mathbf{q}\in\mathcal{P}_n$. If $f\colon(0,\infty)\to\mathbb{R}$ is a convex function, then*

$$0\leq D_f(\mathbf{q},\mathbf{p})-f(1). \tag{4.2}$$

Remark 3 If f is normalized, i.e., $f(1)=0$, then from (4.2) it follows that

$$0\leq D_f(\mathbf{q},\mathbf{p})\quad\text{with}\quad D_f(\mathbf{q},\mathbf{p})=0\quad\text{if and only if}\quad\mathbf{q}=\mathbf{p}. \tag{4.3}$$

Two distributions $\mathbf{q}$ and $\mathbf{p}$ are very similar if $D_f(\mathbf{q},\mathbf{p})$ is very close to zero.

Recently, in [10] a new concept of f-divergences was introduced: when (4.1) is defined for a strongly convex function f, it is denoted with $\tilde{D}_f(\mathbf{q},\mathbf{p})$ and is referred to as strong

f-divergence. Accordingly, in [10] the following improvement of the Csiszár–Körner inequality for strong f-divergences was obtained.

Theorem 8 *Let* $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$. *If* $f : (0,\infty) \to \mathbb{R}$ *is a strongly convex function with modulus* $c > 0$, *then*

$$0 \leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) - f(1) - c\tilde{D}_{\varkappa^2}(\mathbf{q}, \mathbf{p}), \tag{4.4}$$

where $\tilde{D}_{\varkappa^2}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i}\right)^2 - 1$.

Remark 4 Here $\tilde{D}_{\varkappa^2}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i}\right)^2 - 1$ denotes the strong chi-squared distance obtained for the strongly convex function $f(x) = (x-1)^2$ with modulus $c = 1$.

Additionally, if $f(1) = 0$, then from (4.4) we have

$$0 \leq c\tilde{D}_{\varkappa^2}(\mathbf{q}, \mathbf{p}) \leq \tilde{D}_f(\mathbf{q}, \mathbf{p}). \tag{4.5}$$

Inequalities (4.4) and (4.5) improve (4.2) and (4.3).

We further use the results from the previous sections to prove new estimates for strong f-divergences.

Corollary 4 *Let* $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$, $r_i = 1 - \lambda_i \left(1 - \frac{q_i}{p_i}\right)$, *and* $\lambda_i \in [0,1]$, $i \in \{1, \dots, n\}$. *Let* $f : (0,\infty) \to \mathbb{R}$ *be a strongly convex function with modulus* $c > 0$. *Then*

$$\begin{aligned}
0 &\leq \left| \sum_{i=1}^{n} p_i \left| f\left(\frac{q_i}{p_i}\right) - f(r_i) - c(1-\lambda_i)\left(1 - \frac{q_i}{p_i}\right)^2 \right| - \sum_{i=1}^{n} (1-\lambda_i) \left|f'(r_i)\right| |p_i - q_i| \right| \\
&\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) - \sum_{i=1}^{n} p_i f(r_i) \\
&\quad - \sum_{i=1}^{n} (1-\lambda_i) f'(r_i)(p_i - q_i) - c\sum_{i=1}^{n} (1-\lambda_i)^2 (p_i - q_i)^2.
\end{aligned} \tag{4.6}$$

In particular, we have

$$\begin{aligned}
0 &\leq \left| \sum_{i=1}^{n} p_i \left| f\left(\frac{q_i}{p_i}\right) - f(1) - c\left(\frac{q_i}{p_i} - 1\right)^2 \right| - \left|f'(1)\right| \cdot \sum_{i=1}^{n} |p_i - q_i| \right| \\
&\leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) - f(1) - c\tilde{D}_{\chi^2}(\mathbf{q}, \mathbf{p}),
\end{aligned} \tag{4.7}$$

where $\tilde{D}_{\varkappa^2}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i}\right)^2 - 1$.

If, in addition, f *is normalized, then*

$$0 \leq \sum_{i=1}^{n} p_i \left| f\left(\frac{q_i}{p_i}\right) - c\left(\frac{q_i}{p_i} - 1\right)^2 \right| \leq \tilde{D}_f(\mathbf{q}, \mathbf{p}) - c\tilde{D}_{\chi^2}(\mathbf{q}, \mathbf{p}). \tag{4.8}$$

Proof Applying (2.4) to $x_i = \frac{q_i}{p_i}$, $a_i = p_i$ with $\bar{x} = \frac{1}{A_n}\sum_{i=1}^n a_i x_i = \sum_{i=1}^n q_i = 1$ and $\hat{x}_i = (1 - \lambda_i)\bar{x} + \lambda_i x_i = (1-\lambda_i) + \lambda_i \frac{q_i}{p_i} = 1 - \lambda_i\left(1 - \frac{q_i}{p_i}\right) = r_i$, $i \in \{1,\ldots,n\}$, we get

$$\begin{aligned}0 &\le \left|\sum_{i=1}^n p_i \left| f\left(\frac{q_i}{p_i}\right) - f(r_i) - c(1-\lambda_i)\left(1-\frac{q_i}{p_i}\right)^2\right| - \sum_{i=1}^n p_i(1-\lambda_i)\left|f'(r_i)\right|\left|1-\frac{q_i}{p_i}\right|\right| \\ &\le \sum_{i=1}^n p_i f\left(\frac{q_i}{p_i}\right) - \sum_{i=1}^n p_i f(r_i) \\ &- \sum_{i=1}^n p_i(1-\lambda_i) f'(r_i)\left(1-\frac{q_i}{p_i}\right) - c\sum_{i=1}^n p_i(1-\lambda_i)^2\left(1-\frac{q_i}{p_i}\right)^2,\end{aligned}$$

which is equivalent to (4.6).

If $\lambda_i = 0$, $i = 1,\ldots,n$, then $r_i = 1$, $i = 1,\ldots,n$, and from (4.6) we get (4.7). If, in addition, f is normalized, i.e., $f(1) = 0$, then (4.7) implies (4.8). □

Corollary 5 *Let $\lambda_i \in [0,1]$, $i \in \{1,\ldots,n\}$, and let $\mathbf{p},\mathbf{q} \in \mathcal{P}_n$. Suppose $f\colon (0,\infty) \to \mathbb{R}$ is a strongly convex function with modulus $c > 0$. Then*

$$\begin{aligned}&\tilde{D}_f(\mathbf{q},\mathbf{p}) - \sum_{i=1}^n p_i f\left(1-\lambda_i\left(1-\frac{q_i}{p_i}\right)\right) \\ &\le \sum_{i=1}^n (1-\lambda_i) f'\left(\frac{q_i}{p_i}\right)(q_i - p_i) - c\sum_{i=1}^n p_i(1-\lambda_i)^2\left(1-\frac{q_i}{p_i}\right)^2.\end{aligned} \tag{4.9}$$

In particular,

$$\tilde{D}_f(\mathbf{q},\mathbf{p}) - f(1) \le \sum_{i=1}^n f'\left(\frac{q_i}{p_i}\right)(q_i - p_i) - c\tilde{D}_{\chi^2}(\mathbf{q},\mathbf{p}). \tag{4.10}$$

If, in addition, f is normalized, then

$$0 \le \tilde{D}_f(\mathbf{q},\mathbf{p}) \le \sum_{i=1}^n f'\left(\frac{q_i}{p_i}\right)(q_i - p_i) - c\tilde{D}_{\chi^2}(\mathbf{q},\mathbf{p}). \tag{4.11}$$

Proof Applying (2.11) to $x_i = \frac{q_i}{p_i}$ and $a_i = p_i$ with $\bar{x} = \sum_{i=1}^n a_i x_i = \sum_{i=1}^n q_i = 1$, we get

$$\begin{aligned}&\sum_{i=1}^n p_i f\left(\frac{q_i}{p_i}\right) - \sum_{i=1}^n p_i f\left(1-\lambda_i+\lambda_i\frac{q_i}{p_i}\right) \\ &\le \sum_{i=1}^n p_i(1-\lambda_i) f'\left(\frac{q_i}{p_i}\right)\left(\frac{q_i}{p_i}-1\right) - c\sum_{i=1}^n p_i(1-\lambda_i)^2\left(1-\frac{q_i}{p_i}\right)^2,\end{aligned}$$

which is equivalent to (4.9).

Choosing $\lambda_i = 0$, $i = 1,\ldots,n$, from (4.9) we get (4.10). Further, for a normalized function f, (4.10) implies (4.11). □

Corollary 6 *Let $f:(0,\infty)\to\mathbb{R}$ be a strongly convex function with modulus $c>0$. Let $\mathbf{p},\mathbf{q}\in\mathcal{P}_n$ with $\frac{q_i}{p_i}\in[m,M]$, $0<m<M$, and $\lambda_i\in[0,1]$, $i\in\{1,\dots,n\}$. Then*

$$\begin{aligned}
&f(d)+f'(d)(m+M-d-1)+c\sum_{i=1}^{n}p_i\left(m+M-d-\frac{q_i}{p_i}\right)^2\\
&+2c(M-m)^2\sum_{i=1}^{n}\lambda_i(1-\lambda_i)\\
&\le f(m)+f(M)-\tilde{D}_f(\mathbf{q},\mathbf{p})\\
&\le f(e)+f'(m)(m-e)+f'(M)(M-e)-\sum_{i=1}^{n}p_if'\left(\frac{q_i}{p_i}\right)\left(\frac{q_i}{p_i}-e\right)\\
&-c\sum_{i=1}^{n}p_i\left(M-\frac{q_i}{p_i}\right)\left(3\frac{q_i}{p_i}-2e-M\right)+c(m-e)^2
\end{aligned}\tag{4.12}$$

for all $d,e\in[m,M]$.

In particular,

$$\begin{aligned}
&f(m+M-1)+c\tilde{D}_{\varkappa^2}(\mathbf{q},\mathbf{p})+2c(M-m)^2\sum_{i=1}^{n}\lambda_i(1-\lambda_i)\\
&\le f(m)+f(M)-\tilde{D}_f(\mathbf{q},\mathbf{p})\\
&\le f(1)+f'(m)(m-1)+f'(M)(M-1)-\sum_{i=1}^{n}f'\left(\frac{q_i}{p_i}\right)(q_i-p_i)\\
&-c\sum_{i=1}^{n}p_i\left(M-\frac{q_i}{p_i}\right)\left(3\frac{q_i}{p_i}-2-M\right)+c(m-1)^2.
\end{aligned}\tag{4.13}$$

If, in addition, f is normalized, then

$$\begin{aligned}
&f(m+M-1)+c\tilde{D}_f(\mathbf{q},\mathbf{p})+2c(M-m)^2\sum_{i=1}^{n}\lambda_i(1-\lambda_i)\\
&\le f(m)+f(M)-\tilde{D}_f(\mathbf{q},\mathbf{p})\\
&\le f'(m)(m-1)+f'(M)(M-1)-\sum_{i=1}^{n}f'\left(\frac{q_i}{p_i}\right)(q_i-p_i)\\
&-c\sum_{i=1}^{n}p_i\left(M-\frac{q_i}{p_i}\right)\left(3\frac{q_i}{p_i}-2-M\right)+c(m-1)^2.
\end{aligned}\tag{4.14}$$

Proof Applying (3.1) to $x_i=\frac{q_i}{p_i}$ and $a_i=p_i$ with $\bar{x}=\frac{1}{A_n}\sum_{i=1}^{n}a_ix_i=\sum_{i=1}^{n}q_i=1$, we get (4.12). In a particular case, for $e=1$ and $d=m+M-1$, from (4.12) we get (4.13). If, in addition, $f(1)=0$, then (4.13) implies (4.14). □

Applying the previous corollaries to the corresponding generating strongly convex function f, we derive new estimates for some well-known divergences, which are particular cases of the strong f-divergence. Here we consider a few of the most commonly used divergences.

Example 1 The strong Kullback–Leibler divergence of $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$ is defined by

$$\tilde{D}_{KL}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{n} q_i \ln\left(\frac{q_i}{p_i}\right), \tag{4.15}$$

where the generating function is $f(t) = t \ln t$ for $t \in (0, \infty)$. Fix $l > 0$. Since $f''(t) = \frac{1}{t}$, we have $f'' \geq \frac{1}{l}$ on $[m, l]$, $0 < m < l$, and the function $f|_{[m,l]}$ is strongly convex with modulus $c = \frac{1}{2l}$.

Applying inequalities (4.6), (4.8), (4.9), (4.11), (4.12), and (4.14) to $f(t) = t \ln t$ with $c = \frac{1}{2l}$, we may derive new estimates for the strong Kullback–Leibler divergence $\tilde{D}_{KL}(\mathbf{q}, \mathbf{p})$.

Example 2 The strong squared Hellinger divergence of $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$ is defined by

$$\tilde{D}_{h^2}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2,$$

where the generating function is $f(t) = \left(\sqrt{t} - 1\right)^2$ for $t \in (0, \infty)$. Fix $l > 0$. Since $f''(t) = \frac{1}{2\sqrt{t^3}}$, we have $f'' \geq \frac{1}{2\sqrt{l^3}}$ on $[m, l]$, $0 < m < l$, and the function $f|_{[m,l]}$ is strongly convex with modulus $c = \frac{1}{4\sqrt{l^3}}$.

Applying inequalities (4.6), (4.8), (4.9), (4.11), (4.12), and (4.14) to $f(t) = \left(\sqrt{t} - 1\right)^2$ with $c = \frac{1}{4\sqrt{l^3}}$, we may derive new estimates for the strong squared Hellinger divergence $\tilde{D}_{h^2}(\mathbf{q}, \mathbf{p})$.

Example 3 The strong Bhattacharya distance of $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$ is defined by

$$\tilde{D}_{B}(\mathbf{q}, \mathbf{p}) = -\sum_{i=1}^{n} \sqrt{p_i q_i},$$

where the generating function is $f(t) = -\sqrt{t}$ for $t \in (0, \infty)$. Fix $l > 0$. Since $f''(t) = \frac{1}{4\sqrt{t^3}}$, we have $f'' \geq \frac{1}{4\sqrt{l^3}}$ on $[m, l]$, $0 < m < l$, and the function $f|_{[m,l]}$ is strongly convex with modulus $c = \frac{1}{8\sqrt{l^3}}$.

Applying inequalities (4.6), (4.7), (4.8), (4.9), (4.10), (4.11), and (4.12) to $f(t) = -\sqrt{t}$ with $c = \frac{1}{8\sqrt{l^3}}$, we may derive new estimates for the strong Bhattacharya distance $\tilde{D}_{B}(\mathbf{q}, \mathbf{p})$.

Example 4 The strong Jeffreys distance of $\mathbf{p}, \mathbf{q} \in \mathcal{P}_n$ is defined by

$$\tilde{D}_{J}(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{n} (q_i - p_i) \ln \frac{q_i}{p_i} = \tilde{D}_{KL}(\mathbf{q}, \mathbf{p}) + \tilde{D}_{KL}(\mathbf{p}, \mathbf{q}),$$

where the generating function is $f(t) = (t-1)\ln t$ for $t \in (0, \infty)$. Fix $l > 0$. Since $f''(t) = \frac{t+1}{t^2}$, we have $f'' \geq \frac{l+1}{l^2}$ on $[m, l]$, $0 < m < l$, and the function $f|_{[m,l]}$ is strongly convex with modulus $c = \frac{l+1}{2l^2}$.

Applying inequalities (4.6), (4.8), (4.9), (4.11), (4.12), and (4.14) to $f(t) = (t-1)\ln t$ with $c = \frac{l+1}{2l^2}$, we may derive new estimates for the strong Jeffreys distance $\tilde{D}_{J}(\mathbf{q}, \mathbf{p})$.

Example 5 The strong Jensen–Shannon divergence of $\mathbf{p},\mathbf{q}\in\mathcal{P}_n$ is defined by

$$\begin{aligned}\tilde{D}_{JS}(\mathbf{q},\mathbf{p}) &= \frac{1}{2}\left[\sum_{i=1}^{n} q_i \ln\frac{2q_i}{p_i+q_i} + \sum_{i=1}^{n} p_i \frac{2p_i}{p_i+q_i}\right]\\ &= \frac{1}{2}\left[\tilde{D}_{KL}\left(\mathbf{q},\frac{\mathbf{p}+\mathbf{q}}{2}\right) + \tilde{D}_{KL}\left(\mathbf{p},\frac{\mathbf{p}+\mathbf{q}}{2}\right)\right],\end{aligned}$$

where the generating function is $f(t)=\frac{1}{2}\left(t\ln\frac{2t}{1+t}+\ln\frac{2}{1+t}\right)$ for $t\in(0,\infty)$. Fix $l>0$. Since $f''(t)=\frac{1}{2t(1+t)}$, we have $f''\geq\frac{1}{2l(1+l)}$ on $[m,l]$, $0<m<l$, and the function $f|_{[m,l]}$ is strongly convex with modulus $c=\frac{1}{4l(1+l)}$.

Applying inequalities (4.6)), (4.8), (4.9), (4.11), (4.12), and (4.14) to $f(t)=\frac{1}{2}\left(t\ln\frac{2t}{1+t}+\ln\frac{2}{1+t}\right)$ with $c=\frac{1}{4l(1+l)}$, we may derive new estimates for the strong Jensen–Shannon divergence $\tilde{D}_{JS}(\mathbf{q},\mathbf{p})$.

We now consider the Shannon entropy [25], defined for a random variable X in terms of its probability distribution $\mathbf{p}$ as

$$S(\mathbf{p})=\sum_{i=1}^{n} p_i\ln\frac{1}{p_i} = -\sum_{i=1}^{n} p_i\ln p_i. \tag{4.16}$$

It quantifies the unevenness in $\mathbf{p}$ and satisfies the relation

$$0\leq S(\mathbf{p})\leq \ln n.$$

Using the results from the previous sections, we obtain new estimates for the Shannon entropy.

Corollary 7 *Let $l>0$, and let $\mathbf{p}\in\mathcal{P}_n$ be such that $\frac{1}{p_1},\dots,\frac{1}{p_n}\in(0,l]$. Let $\bar{p}_i=n-\lambda_i\left(n-\frac{1}{p_i}\right)$, $\lambda_i\in[0,1]$, $i\in\{1,\dots,n\}$. Then*

$$\begin{aligned}S(\mathbf{p}) &\leq S(\mathbf{p}) + \left|\sum_{i=1}^{n} p_i\left|\ln p_i\bar{p}_i - \frac{1-\lambda_i}{2l^2}\left(n-\frac{1}{p_i}\right)^2\right| - \sum_{i=1}^{n}\frac{p_i}{\bar{p}_i}(1-\lambda_i)\left|n-\frac{1}{p_i}\right|\right|\\ &\leq \sum_{i=1}^{n} p_i\ln\bar{p}_i + \sum_{i=1}^{n}\frac{p_i}{\bar{p}_i}(1-\lambda_i)\left(n-\frac{1}{p_i}\right) - \frac{1}{2l^2}\sum_{i=1}^{n} p_i(1-\lambda_i)^2\left(n-\frac{1}{p_i}\right)^2.\end{aligned} \tag{4.17}$$

In particular, we have

$$\begin{aligned}S(\mathbf{p}) &\leq S(\mathbf{p}) + \left|\sum_{i=1}^{n} p_i\left|\ln p_i\bar{p}_i - \frac{1}{2l^2}\left(n-\frac{1}{p_i}\right)^2\right| - \sum_{i=1}^{n}\frac{p_i}{\bar{p}_i}\left|n-\frac{1}{p_i}\right|\right|\\ &\leq \sum_{i=1}^{n} p_i\ln\bar{p}_i + \sum_{i=1}^{n}\frac{p_i}{\bar{p}_i}\left(n-\frac{1}{p_i}\right) - \frac{1}{2l^2}\sum_{i=1}^{n} p_i\left(n-\frac{1}{p_i}\right)^2.\end{aligned} \tag{4.18}$$

Proof Applying (2.4) to the function $f(t)=-\ln t$, $t\in(0,l]$, strongly convex with modulus $c=\frac{1}{2l^2}$, and $x_i=\frac{1}{p_i}$ and $a_i=p_i$ with $\bar{x}=\frac{1}{A_n}\sum_{i=1}^{n} a_i x_i=\sum_{i=1}^{n} p_i\frac{1}{p_i}=n$ and $\hat{x}_i=(1-\lambda_i)\bar{x}+\lambda_i x_i=$

$(1-\lambda_i)n+\lambda_i\frac{1}{p_i}=n-\lambda_i\left(n-\frac{1}{p_i}\right)=\bar{p}_i$, $i\in\{1,\ldots,n\}$, we get

$$\begin{aligned}0&\le\left|\sum_{i=1}^{n}p_i\left|-\ln\frac{1}{p_i}+\ln\bar{p}_i-\frac{1-\lambda_i}{2l^2}\left(n-\frac{1}{p_i}\right)^2\right|-\sum_{i=1}^{n}p_i(1-\lambda_i)\left|\frac{1}{\bar{p}_i}\right|\left|n-\frac{1}{p_i}\right|\right|\\&\le-\sum_{i=1}^{n}p_i\ln\frac{1}{p_i}+\sum_{i=1}^{n}p_i\ln\bar{p}_i+\sum_{i=1}^{n}\frac{p_i}{\bar{p}_i}(1-\lambda_i)\left(n-\frac{1}{p_i}\right)\\&-\frac{1}{2l^2}\sum_{i=1}^{n}p_i(1-\lambda_i)^2\left(n-\frac{1}{p_i}\right)^2,\end{aligned}$$

which is equivalent to (4.17).

Choosing $\lambda_i=0$, $i=1,\ldots,n$, from (4.17) we get (4.18). □

Corollary 8 *Let $l>0$, let $\mathbf{p}\in\mathcal{P}_n$ be such that $\frac{1}{p_1},\ldots,\frac{1}{p_n}\in(0,l]$, and let $\lambda_i\in[0,1]$, $i\in\{1,\ldots,n\}$. Then*

$$\begin{aligned}&\sum_{i=1}^{n}p_i^2(1-\lambda_i)\left(\frac{1}{p_i}-n\right)+\frac{1}{2l^2}\sum_{i=1}^{n}p_i(1-\lambda_i)^2\left(\frac{1}{p_i}-n\right)^2\\&+\sum_{i=1}^{n}p_i\ln\left((1-\lambda_i)n+\frac{\lambda_i}{p_i}\right)\\&\le S(\mathbf{p}).\end{aligned}\tag{4.19}$$

In particular, we have

$$\ln n+1-n\sum_{i=1}^{n}p_i^2+\frac{1}{2l^2}\sum_{i=1}^{n}p_i\left(\frac{1}{p_i}-n\right)^2\le S(\mathbf{p}).\tag{4.20}$$

Proof Applying (2.11) to the strongly convex function $f(t)=-\ln t$, $t\in(0,l]$, with modulus $c=\frac{1}{2l^2}$, and to $x_i=\frac{1}{p_i}$ and $a_i=p_i$ with $\bar{x}=\frac{1}{A_n}\sum_{i=1}^{n}a_ix_i=\sum_{i=1}^{n}p_i\frac{1}{p_i}=n$, we get

$$\begin{aligned}&-\sum_{i=1}^{n}p_i\ln\frac{1}{p_i}+\sum_{i=1}^{n}p_i\ln\left((1-\lambda_i)n+\frac{\lambda_i}{p_i}\right)\\&\le-\sum_{i=1}^{n}p_i(1-\lambda_i)\left(\frac{1}{p_i}\right)^{-1}\left(\frac{1}{p_i}-n\right)-\frac{1}{2l^2}\sum_{i=1}^{n}p_i(1-\lambda_i)^2\left(\frac{1}{p_i}-n\right)^2,\end{aligned}$$

which is equivalent to (4.19). If we choose $\lambda_i=0$, $i=1,\ldots,n$, then (4.19) implies (4.20). □

Corollary 9 *Let $0<m<l$, let $\mathbf{p}\in\mathcal{P}_n$ be such that $\frac{1}{p_1},\ldots,\frac{1}{p_n}\in[m,l]$, and let $\lambda_i\in[0,1]$, $i\in\{1,\ldots,n\}$. Then*

$$\begin{aligned}&\frac{1}{2l^2}\sum_{i=1}^{n}p_i\left(m+l-d-\frac{1}{p_i}\right)^2-\frac{1}{d}(m+l-d-n)\\&+\frac{(l-m)^2}{l^2}\sum_{i=1}^{n}\lambda_i(1-\lambda_i)+\ln\frac{ml}{d}\end{aligned}\tag{4.21}$$

$$
\begin{aligned}
&\leq S(\mathbf{p}) \\
&\leq \frac{1}{m}(e-m) + \frac{1}{l}(e-l) + \sum_{i=1}^{n} p_i^2 \left(e - \frac{1}{p_i}\right) + \ln \frac{ml}{e} \\
&- \frac{1}{2l^2} \sum_{i=1}^{n} p_i \left(l - \frac{1}{p_i}\right)\left(\frac{3}{p_i} - 2e - l\right) + \frac{(m-e)^2}{2l^2}
\end{aligned}
$$

for all $d, e \in [m, l]$.

In particular, we have

$$
\begin{aligned}
&\frac{1}{2l^2} \sum_{i=1}^{n} p_i \left(n - \frac{1}{p_i}\right)^2 + \frac{(l-m)^2}{l^2} \sum_{i=1}^{n} \lambda_i (1-\lambda_i) + \ln \frac{ml}{m+l-n} \\
&\leq S(\mathbf{p}) \\
&\leq \frac{1}{m}(n-m) + \frac{1}{l}(n-l) + \sum_{i=1}^{n} p_i^2 \left(n - \frac{1}{p_i}\right) + \ln \frac{ml}{n} \\
&- \frac{1}{2l^2} \sum_{i=1}^{n} p_i \left(l - \frac{1}{p_i}\right)\left(\frac{3}{p_i} - 2n - l\right) + \frac{(m-n)^2}{2l^2}.
\end{aligned}
\tag{4.22}
$$

Proof Applying (3.1) to the strongly convex function $f(t) = -\ln t$, $t \in (0, l]$, with modulus $c = \frac{1}{2l^2}$ and to $x_i = \frac{1}{p_i}$ and $a_i = p_i$ with $\bar{x} = \frac{1}{A_n} \sum_{i=1}^{n} a_i x_i = \sum_{i=1}^{n} p_i \frac{1}{p_i} = n$, we get

$$
\begin{aligned}
&-\ln d - \frac{1}{d}(m+l-d-n) + \frac{1}{2l^2} \sum_{i=1}^{n} p_i \left(m+l-d-\frac{1}{p_i}\right)^2 \\
&+ \frac{(l-m)^2}{l^2} \sum_{i=1}^{n} \lambda_i (1-\lambda_i) \\
&\leq -\ln m - \ln l + \sum_{i=1}^{n} p_i \ln \frac{1}{p_i} \\
&\leq -\ln e - \frac{1}{m}(m-e) - \frac{1}{l}(l-e) + \sum_{i=1}^{n} p_i \left(\frac{1}{p_i}\right)^{-1} \left(\frac{1}{p_i} - e\right) \\
&- \frac{1}{2l^2} \sum_{i=1}^{n} p_i \left(l - \frac{1}{p_i}\right)\left(l - \frac{3}{p_i} + 2e\right) + \frac{(m-e)^2}{2l^2},
\end{aligned}
$$

which is equivalent to (4.21). Choosing $e = n$ and $d = m + l - n$, from (4.21) we get (4.22). □

5 New bounds for the Chebyshev functional

One of the fundamental inequalities in probability is the discrete Chebyshev inequality, which we quote in the following form (see [21]).

Theorem 9 *Let* $\boldsymbol{a} = (a_1, \ldots, a_n)$ *be a nonnegative n-tuple with* $A_n = \sum_{i=1}^n a_i > 0$, *and let* $\boldsymbol{p} = (p_1, \ldots, p_n)$ *and* $\boldsymbol{q} = (q_1, \ldots, q_n)$ *be monotonic real n-tuples in the same direction. Then*

$$\frac{1}{A_n}\sum_{i=1}^n a_i p_i q_i - \frac{1}{A_n^2}\sum_{i=1}^n a_i p_i \sum_{i=1}^n a_i q_i \geq 0. \tag{5.1}$$

If $\boldsymbol{p}$ *and* $\boldsymbol{q}$ *are monotonic in the opposite direction, then we have the reverse inequality of* (5.1).

We can find many papers that study the Chebyshev functional $T(\boldsymbol{a};\boldsymbol{p},\boldsymbol{q})$ derived from the Chebyshev inequality (5.1) by subtracting its right side from its left one:

$$T(\boldsymbol{a};\boldsymbol{p},\boldsymbol{q}) = A_n \sum_{i=1}^n a_i p_i q_i - \sum_{i=1}^n a_i p_i \sum_{i=1}^n a_i q_i, \tag{5.2}$$

and in the normalized form as

$$\bar{T}(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{n}\sum_{i=1}^n p_i q_i - \frac{1}{n^2}\sum_{i=1}^n p_i \sum_{i=1}^n q_i. \tag{5.3}$$

By (5.1) we have

$$T(\boldsymbol{a};\boldsymbol{p},\boldsymbol{q}) \geq 0 \quad \text{and} \quad \bar{T}(\boldsymbol{p},\boldsymbol{q}) \geq 0.$$

Using the results from Sect. 2, we obtain improvements of the Chebyshev inequality (5.1), i.e., we get new bounds for the Chebishev functional of types (5.2) and (5.3) without the assumption of monotonicity.

Corollary 10 *Let* $\boldsymbol{a} = (a_1, \ldots, a_n)$ *be a nonnegative n-tuple with* $A_n = \sum_{i=1}^n a_i > 0$, *and let* $\boldsymbol{p} = (p_1, \ldots, p_n)$ *and* $\boldsymbol{q} = (q_1, \ldots, q_n)$ *be real n-tuples with* $\bar{p} = \frac{1}{A_n}\sum_{i=1}^n a_i p_i$ *and* $P_n = \sum_{i=1}^n p_i$. *Then*

$$\begin{aligned}
0 &\leq \frac{c}{A_n}\sum_{i=1}^n a_i (p_i - \bar{p})^2 \\
&\leq \left| \frac{1}{A_n}\sum_{i=1}^n a_i \left| f(p_i) - f(\bar{p}) - c(p_i - \bar{p})^2 \right| - \left| f'(\bar{p}) \right| \cdot \frac{1}{A_n}\sum_{i=1}^n a_i \left| (p_i - \bar{p}) \right| \right| \\
&+ \frac{c}{A_n}\sum_{i=1}^n a_i (p_i - \bar{p})^2 \\
&\leq \frac{1}{A_n}\sum_{i=1}^n a_i f(p_i) - f(\bar{p}) \\
&\leq T(\boldsymbol{a};\boldsymbol{p},\boldsymbol{q}) - \frac{c}{A_n}\sum_{i=1}^n a_i (\bar{p} - p_i)^2 \leq T(\boldsymbol{a};\boldsymbol{p},\boldsymbol{q}).
\end{aligned} \tag{5.4}$$

In particular, we have

$$\begin{aligned}
0 &\le \frac{c}{n}\sum_{i=1}^{n}\left(p_i-\frac{P_n}{n}\right)^2 \qquad (5.5)\\
&\le \left|\frac{1}{A_n}\sum_{i=1}^{n}a_i\left|f(p_i)-f\left(\frac{P_n}{n}\right)-c\left(p_i-\frac{P_n}{n}\right)^2\right|-\left|f'\left(\frac{P_n}{n}\right)\right|\cdot\frac{1}{n}\sum_{i=1}^{n}\left|\left(p_i-\frac{P_n}{n}\right)\right|\right|\\
&+\frac{c}{n}\sum_{i=1}^{n}\left(p_i-\frac{P_n}{n}\right)^2\\
&\le \frac{1}{n}\sum_{i=1}^{n}f(p_i)-f\left(\frac{P_n}{n}\right)\\
&\le T(\boldsymbol{p},\boldsymbol{q})-\frac{c}{n}\sum_{i=1}^{n}\left(p_i-\frac{P_n}{n}\right)^2\le T(\boldsymbol{p},\boldsymbol{q}).
\end{aligned}$$

Proof Combining inequalities (2.8) and (2.12), we have

$$\begin{aligned}
0 &\le \frac{c}{A_n}\sum_{i=1}^{n}a_i(x_i-\bar{x})^2 \qquad (5.6)\\
&\le \left|\frac{1}{A_n}\sum_{i=1}^{n}a_i\left|f(x_i)-f(\bar{x})-c(x_i-\bar{x})^2\right|-\left|f'(\bar{x})\right|\cdot\frac{1}{A_n}\sum_{i=1}^{n}a_i\left|(x_i-\bar{x})\right|\right|\\
&+\frac{c}{A_n}\sum_{i=1}^{n}a_i(x_i-\bar{x})^2\\
&\le \frac{1}{A_n}\sum_{i=1}^{n}a_if(x_i)-f(\bar{x})\\
&\le \frac{1}{A_n}\sum_{i=1}^{n}a_if'(x_i)x_i-\frac{1}{A_n^2}\sum_{i=1}^{n}a_ix_i\sum_{i=1}^{n}a_if'(x_i)-\frac{c}{A_n}\sum_{i=1}^{n}a_i(\bar{x}-x_i)^2\\
&\le \frac{1}{A_n}\sum_{i=1}^{n}a_if'(x_i)x_i-\frac{1}{A_n^2}\sum_{i=1}^{n}a_ix_i\sum_{i=1}^{n}a_if'(x_i).
\end{aligned}$$

Setting $f'(x_i)=q_i$ and $x_i=p_i$, $i\in\{1,\dots,n\}$ and using (5.6), we get (5.4).

If we set $a_i=\frac{1}{n}$, $i=1,\dots,n$, then $\bar{p}=\frac{1}{n}\sum_{i=1}^{n}p_i=\frac{P_n}{n}$, where $P_n=\sum_{i=1}^{n}p_i$. Now inequality (5.5) immediately follows from (5.4). □

Author contributions
Each author contributed the same level of work.

Funding
This research is partially supported through KK.01.1.1.02.0027, a project cofinanced by the Croatian Government and the European Union through the European Regional Development Fund – the Competitiveness and Cohesion Operational Programme.

Data Availability
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Received: 20 March 2024 Accepted: 13 August 2024 Published online: 02 September 2024

References

1. Adil Khan, M., Husain, Z., Chu, Y.-M.: New estimates for Csiszár divergence and Zipf–Mandelbrot entropy via Jensen–Mercer's inequality. Complexity **2020**, 1–8 (2020)
2. Butt, S.I., Agarwal, P., Yousaf, S., Guirao, J.L.G.: Generalized fractal Jensen and Jensen–Mercer inequalities for harmonic convex function with applications. J. Inequal. Appl. **2022**, 1 (2022)
3. Butt, S.I., Sayyari, Y., Agarwal, P., Nieto, J.J., Umar, M.: On some inequalities for uniformly convex mapping with estimations to normal distributions. J. Inequal. Appl. **2023**, 89 (2023)
4. Crooks, G.E.: On measures of entropy and information. Tech. Note 009 v0.8 (2021)
5. Csiszár, I.: Information-type measures of difference of probability functions and indirect observations. Studia Sci. Math. Hung. **2**, 299–318 (1967)
6. Csiszár, I., Körner, J.: Information Theory: Coding Theorem for Discrete Memoryless Systems. Academic Press, New York (1981)
7. Dragomir, S.S., Ionescu, N.M.: Some converse of Jensen's inequality and applications. Rev. Anal. Numér. Théor. Approx. **23**, 71–78 (1994)
8. Dragomir, S.S., Scarmozzino, F.P.: A Refinement of Jensen's discrete inequality for differentiable convex functions. RGMIA Res. Rep. Collect. **5**(4) (2002)
9. Horváth, L.: Some notes on Jensen–Mercer's type inequalities; extensions and refinements with applications. Math. Inequal. Appl. **24**(4), 1093–1111 (2021)
10. Ivelić Bradanović, S.: Sherman's inequality and its converse for strongly convex functions with applications to generalized f-divergences. Turk. J. Math. **6**(43), 2680–2696 (2019)
11. Ivelić Bradanović, S.: Improvements of Jensen's inequality and its converse for strongly convex functions with applications to strongly f-divergences. J. Math. Anal. Appl. **2**(531), 1–16 (2024)
12. Ivelić, S., Matković, A., Pečarić, J.: On a Jensen–Mercer operator inequality. Banach J. Math. Anal. **5**(1), 19–28 (2011)
13. Jarad, F., Sahoo, S.K., Nisar, K.S., Treanţă, S., Emadifar, H., Botmart, T.: New stochastic fractional integral and related inequalities of Jensen–Mercer and Hermite–Hadamard–Mercer type for convex stochastic processes. J. Inequal. Appl. **2023**, 51 (2023)
14. Khan, A.R., Rubab, F.: Mercer type variants of the Jensen–Steffensen inequality. Rocky Mt. J. Math. **52**(5), 1693–1712 (2022)
15. Klaričić Bakula, M., Matić, M., Pečarić, J.: On some general inequalities related to Jensen's inequality. Int. Ser. Numer. Math. **157**, 233–243 (2008)
16. Krnić, M., Lovričević, N., Pečarić, J.: On some properties of Jensen–Mercer's functional. J. Math. Inequal. **6**(1), 125–139 (2012)
17. Mercer, A.M.: A variant of Jensen's inequality. JIPAM. J. Inequal. Pure Appl. Math. **4**(4), 1–2 (2003)
18. Moradi, H.R., Omidvar, M.E., Adil Khan, M., Nikodem, K.: Around Jensen's inequality for strongly convex functions. Aequ. Math. **92**, 25–37 (2018)
19. Niculescu, C.P., Persson, L.E.: Convex Functions and Their Applications. A Contemporary Approach, 2nd edn. CMS Books in Mathematics, vol. 2. Springer, New York (2018)
20. Nikodem, K.: On Strongly Convex Functions and Related Classes of Functions, Handbook of Functional Equations, pp. 365–405. Springer, New York (2014)
21. Pečarić, J., Proschan, F., Tong, Y.L.: Convex Functions, Partial Orderings and Statistical Applications. Academic Press, New York (1992)
22. Polyanskiy, Y., Wu, Y.: Lecture, Information Theory: From Coding to Learning. Cambridge University Press, Cambridge (2022)
23. Roberts, A.W., Varberg, D.E.: Convex Functions. Academic Press, New York (1973)
24. Sayyari, Y., Barsam, H.: Jensen–Mercer inequality for uniformly convex functions with some applications. Afr. Math. **34**, 38 (2023)
25. Shannon, C.E., Weaver, W.: The Mathemtiatical Theory of Comnunication, Urbana. University of Illinois Press, Champaign (1949)

Publisher's Note