

RESEARCH

Open Access



# Refinements of the integral Jensen's inequality generated by finite or infinite permutations

László Horváth<sup>1\*</sup> 

\*Correspondence:  
lhorvath@almos.uni-pannon.hu

<sup>1</sup>Department of Mathematics,  
University of Pannonia, Veszprém,  
Hungary

## Abstract

There are a lot of papers dealing with applications of the so-called cyclic refinement of the discrete Jensen's inequality. A significant generalization of the cyclic refinement, based on combinatorial considerations, has recently been discovered by the author. In the present paper we give the integral versions of these results. On the one hand, a new method to refine the integral Jensen's inequality is developed. On the other hand, the result contains some recent refinements of the integral Jensen's inequality as elementary cases. Finally some applications to the Fejér inequality (especially the Hermite–Hadamard inequality), quasi-arithmetic means, and  $f$ -divergences are presented.

**MSC:** 26D15; 26B25; 94A17

**Keywords:** Discrete and integral Jensen's inequalities; Refinements; Fejér inequality; Quasi-arithmetic means;  $f$ -divergences

## 1 Introduction

The significance of convex functions is rightly due to Jensen's inequality. A real function  $f$  defined on an interval  $C \subset \mathbb{R}$  is called convex if it satisfies

$$f(\alpha t_1 + (1 - \alpha)t_2) \leq \alpha f(t_1) + (1 - \alpha)f(t_2)$$

for all  $t_1, t_2 \in C$  and all  $\alpha \in [0, 1]$ .

Let the set  $I$  denote either  $\{1, \dots, n\}$  for some  $n \geq 1$  or  $\mathbb{N}_+$ . We say that the numbers  $(p_i)_{i \in I}$  represent a discrete probability distribution if  $p_i \geq 0$  ( $i \in I$ ) and  $\sum_{i \in I} p_i = 1$ . It is called positive if  $p_i > 0$  ( $i \in I$ ). A permutation  $\pi$  of  $I$  refers to a bijection from  $I$  onto itself.

The following discrete and integral versions of Jensen's inequality are well known.

**Theorem 1** (discrete Jensen's inequalities, see [16] and [17]) (a) *Let  $C$  be a convex subset of a real vector space  $V$ , and let  $f : C \rightarrow \mathbb{R}$  be a convex function. If  $p_1, \dots, p_n$  represent a*

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

discrete probability distribution and  $v_1, \dots, v_n \in C$ , then

$$f\left(\sum_{i=1}^n p_i v_i\right) \leq \sum_{i=1}^n p_i f(v_i). \tag{1}$$

(b) Let  $C$  be a closed convex subset of a real Banach space  $V$ , and let  $f : C \rightarrow \mathbb{R}$  be a convex function. If  $p_1, p_2, \dots$  represent a discrete probability distribution and  $v_1, v_2, \dots \in C$  such that the series  $\sum_{i=1}^\infty p_i v_i$  and  $\sum_{i=1}^\infty p_i f(v_i)$  are absolutely convergent, then

$$f\left(\sum_{i=1}^\infty p_i v_i\right) \leq \sum_{i=1}^\infty p_i f(v_i). \tag{2}$$

**Theorem 2** (integral Jensen’s inequality, see [16]) *Let  $\varphi$  be an integrable function on a probability space  $(X, \mathcal{A}, \mu)$  taking values in an interval  $C \subset \mathbb{R}$ . Then  $\int_X \varphi d\mu$  lies in  $C$ . If  $f$  is a convex function on  $C$  such that  $f \circ \varphi$  is  $\mu$ -integrable, then*

$$f\left(\int_X \varphi d\mu\right) \leq \int_X f \circ \varphi d\mu. \tag{3}$$

*Remark 3* It follows that Theorem 1 (b) can be generalized in case of  $V = \mathbb{R}$ : if  $C \subset \mathbb{R}$  is an interval (not necessarily closed) and the other conditions of the statement are satisfied, then  $\sum_{i=1}^\infty p_i v_i$  lies in  $C$  and (2) holds.

There are many papers dealing with refinements of discrete and integral Jensen’s inequalities (see the book [12] and the references therein).

In papers [2] and [13] there are special refinements of the discrete Jensen’s inequality of the form in Theorem 1 (a) (so-called cyclic refinements). These led the author to the following refinement of the discrete Jensen’s inequality which is a significant generalization of the previously mentioned results.

**Theorem 4** (see [11]) (a) *Let  $k, n \geq 2$  be integers, and let  $p_1, \dots, p_n$  and  $\lambda_1, \dots, \lambda_k$  represent positive probability distributions. For each  $j = 1, \dots, k$ , let  $\pi_j$  be a permutation of the set  $\{1, \dots, n\}$ . If  $C$  is a convex subset of a real vector space  $V$ ,  $f : C \rightarrow \mathbb{R}$  is a convex function, and  $v_1, \dots, v_n \in C$ , then*

$$\begin{aligned} f\left(\sum_{i=1}^n p_i v_i\right) &\leq C_{per} = C_{per}(f, \mathbf{v}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\pi}) \\ &:= \sum_{i=1}^n \left(\sum_{j=1}^k \lambda_j p_{\pi_j(i)}\right) f\left(\frac{\sum_{j=1}^k \lambda_j p_{\pi_j(i)} v_{\pi_j(i)}}{\sum_{j=1}^k \lambda_j p_{\pi_j(i)}}\right) \leq \sum_{i=1}^n p_i f(v_i). \end{aligned}$$

(b) *Let the set  $J$  denote either  $\{1, \dots, k\}$  for some  $k \geq 2$  or  $\mathbb{N}_+$ . Let  $p_1, p_2, \dots$  and  $(\lambda_j)_{j \in J}$  represent positive probability distributions. For each  $j \in J$ , let  $\pi_j$  be a permutation of the set  $\mathbb{N}_+$ . If  $C$  is a closed convex subset of a real Banach space  $(V, \|\cdot\|)$ ,  $f : C \rightarrow \mathbb{R}$  is a convex function, and  $v_1, v_2, \dots \in C$  such that the series  $\sum_{i=1}^\infty p_i v_i$  and  $\sum_{i=1}^\infty p_i f(v_i)$  are absolutely*

convergent, then

$$\begin{aligned}
 f\left(\sum_{i=1}^{\infty} p_i v_i\right) &\leq C_{per} = C_{per}(f, \mathbf{v}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\pi}) \\
 &:= \sum_{i=1}^{\infty} \left(\sum_{j \in J} \lambda_j p_{\pi_j(i)}\right) f\left(\frac{\sum_{j \in J} \lambda_j p_{\pi_j(i)} v_{\pi_j(i)}}{\sum_{j \in J} \lambda_j p_{\pi_j(i)}}\right) \leq \sum_{i=1}^{\infty} p_i f(v_i).
 \end{aligned}$$

In the paper [13] we obtain refinements of the integral Jensen’s inequality by using cyclic refinements of the discrete Jensen’s inequality, but these results are not natural obverses of the discrete one’s.

In this paper we give the integral version of Theorem 4 when  $V = \mathbb{R}$ . On the one hand, a new method to refine the integral Jensen’s inequality is developed (totally different from earlier techniques, see e.g. [10] and [19]) and our result contains Theorem 4 when  $V = \mathbb{R}$ . On the other hand, we can have from it some recent refinements of the integral Jensen’s inequality (see [7], [5], [6]) as elementary cases. Finally, some applications to the Fejér inequality (especially the Hermite–Hadamard inequality), quasi-arithmetic means, and  $f$ -divergences are presented.

## 2 Preliminary result

We give an extension of Theorem 4 if  $V = \mathbb{R}$ .

**Proposition 5** *Let the index set  $I$  denote either  $\{1, \dots, n\}$  for some  $n \geq 1$  or  $\mathbb{N}_+$ . Let the index set  $J$  denote either  $\{1, \dots, k\}$  for some  $k \geq 1$  or  $\mathbb{N}_+$ . For each  $j \in J$ , let  $\pi_j$  be a permutation of the set  $I$ . Let  $(p_i)_{i \in I}$  and  $(\lambda_j)_{j \in J}$  represent positive probability distributions. If  $C$  is an interval in  $\mathbb{R}$ ,  $f : C \rightarrow \mathbb{R}$  is a convex function, and  $(v_i)_{i \in I}$  is a sequence from  $C$  such that the series  $\sum_{i \in I} p_i v_i$  and  $\sum_{i \in I} p_i f(v_i)$  are absolutely convergent, then*

$$\begin{aligned}
 f\left(\sum_{i \in I} p_i v_i\right) &\leq C_{per} = C_{per}(f, \mathbf{v}, \mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\pi}) \\
 &:= \sum_{i \in I} \left(\sum_{j \in J} \lambda_j p_{\pi_j(i)}\right) f\left(\frac{\sum_{j \in J} \lambda_j p_{\pi_j(i)} v_{\pi_j(i)}}{\sum_{j \in J} \lambda_j p_{\pi_j(i)}}\right) \leq \sum_{i \in I} p_i f(v_i).
 \end{aligned}$$

*Proof* By using Remark 3, we can copy the proof of Theorem 4 in [11]. □

The positive part  $f^+$  and the negative part  $f^-$  of a real-valued function  $f$  are defined in the usual way.

We need another result about integrability.

**Lemma 6** *Let  $\varphi$  be an integrable function on a probability space  $(X, \mathcal{A}, \mu)$  taking values in an interval  $C \subset \mathbb{R}$ . If  $f$  is a convex function on  $C$  such that  $f \circ \varphi$  is  $\mu$ -integrable, then there exists a convex function  $g$  on  $C$  such that  $|f| \leq g$  and  $g \circ \varphi$  is  $\mu$ -integrable too.*

*Proof* Along with the function  $f$ , the function  $f^+$  is also convex. Since  $f \circ \varphi$  is  $\mu$ -integrable,  $f^+ \circ \varphi$  is also  $\mu$ -integrable.

The convexity of  $f$  on  $C$  shows that there is an affine function  $l : C \rightarrow \mathbb{R}$ ,  $l(t) = at + b$  for which

$$f(t) \geq at + b, \quad t \in C.$$

Then

$$\begin{aligned} f^-(t) &= \max(-f(t), 0) \leq \max(-at - b, 0) = l^-(t) \\ &\leq |at + b| = |l(t)|, \quad t \in C. \end{aligned}$$

Since the function  $\varphi$  is a  $\mu$ -integrable function,  $|l| \circ \varphi$  is also  $\mu$ -integrable.

Using that the function  $|l|$  is convex,  $|f| = f^+ + f^-$ , and the sum of two convex functions is also convex, it follows from the above that  $g := f^+ + |l|$  can be chosen.

The proof is complete. □

We shall use the following Fubini theorem for double series.

**Theorem 7** (see [20]) *Let  $a(i, j) \in \mathbb{R}$  ( $(i, j) \in \mathbb{N}_+ \times \mathbb{N}_+$ ). If any of the two sums*

$$\sum_{i=1}^{\infty} \left( \sum_{j=1}^{\infty} |a(i, j)| \right), \quad \sum_{j=1}^{\infty} \left( \sum_{i=1}^{\infty} |a(i, j)| \right)$$

*is finite, then both of the series*

$$\sum_{i=1}^{\infty} \left( \sum_{j=1}^{\infty} a(i, j) \right), \quad \sum_{j=1}^{\infty} \left( \sum_{i=1}^{\infty} a(i, j) \right)$$

*are absolutely convergent and both have the same sum.*

### 3 Main results

We need the following hypotheses.

- (H<sub>1</sub>) Let  $(X, \mathcal{A}, \mu)$  be a probability space.
- (H<sub>2</sub>) Let the index set  $I$  denote either  $\{1, \dots, n\}$  for some  $n \geq 1$  or  $\mathbb{N}_+$ . Let the index set  $J$  denote either  $\{1, \dots, k\}$  for some  $k \geq 1$  or  $\mathbb{N}_+$ .
- (H<sub>3</sub>) Let  $(\lambda_j)_{j \in J}$  represent a positive probability distribution. For each  $j \in J$ , let  $\pi_j$  be a permutation of the set  $I$ .
- (H<sub>4</sub>) Suppose that we are given a sequence  $\mathfrak{M}_I = (\mu_i)_{i \in I}$  of measures on  $\mathcal{A}$  with  $\mu_i(X) > 0$  for all  $i \in I$  and  $\sum_{i \in I} \mu_i = \mu$ .
- (H<sub>5</sub>) Suppose that we are given a sequence  $\mathfrak{S}_I = (A_i)_{i \in I}$  of pairwise disjoint sets  $A_i \in \mathcal{A}$  with  $\mu(A_i) > 0$  for all  $i \in I$  and  $\bigcup_{i \in I} A_i = X$ .

**Theorem 8** *Assume (H<sub>1</sub>–H<sub>3</sub>) and (H<sub>4</sub>). Let  $C \subset \mathbb{R}$  be an interval, and  $f : C \rightarrow \mathbb{R}$  be a convex function. Let  $\varphi$  be a  $\mu$ -integrable function on  $X$  taking values in  $C$  such that  $f \circ \varphi$  is*

also  $\mu$ -integrable on  $X$ . Then

$$\begin{aligned} f\left(\int_X \varphi \, d\mu\right) &\leq C_{mes} = C_{mes}(f, \varphi, \lambda, \pi, \mathfrak{M}_I) \\ &:= \sum_{i \in I} \left(\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}(X)\right) f\left(\frac{\sum_{j \in J} \lambda_j \int_X \varphi \, d\mu_{\pi_j(i)}}{\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}(X)}\right) \leq \int_X f \circ \varphi \, d\mu. \end{aligned}$$

*Proof* This can be obtained by an application of Proposition 5 to the parameters

$$p_i := \mu_i(X) \quad \text{and} \quad v_i := \frac{1}{\mu_i(X)} \int_X \varphi \, d\mu_i, \quad i \in I.$$

Really,  $(p_i)_{i \in I}$  represents a positive probability distribution, and by the integral Jensen’s inequality,  $v_i \in C$  ( $i \in I$ ).

Next we show that the series  $\sum_{i \in I} p_i v_i$  and  $\sum_{i \in I} p_i f(v_i)$  are absolutely convergent. Since  $\varphi$  is a  $\mu$ -integrable function on  $X$  and  $\sum_{i \in I} \mu_i = \mu$ ,

$$\sum_{i \in I} p_i |v_i| = \sum_{i \in I} \left| \int_X \varphi \, d\mu_i \right| \leq \sum_{i \in I} \int_X |\varphi| \, d\mu_i = \int_X |\varphi| \, d\mu < \infty.$$

By Lemma 6, there exists a convex function  $g$  on  $C$  such that

$$|f| \leq g \quad \text{and} \quad g \circ \varphi \text{ is } \mu\text{-integrable.} \tag{4}$$

Another application of the integral Jensen’s inequality and  $\sum_{i \in I} \mu_i = \mu$  now show that

$$\begin{aligned} \sum_{i \in I} p_i |f(v_i)| &= \sum_{i \in I} \mu_i(X) \left| f\left(\frac{1}{\mu_i(X)} \int_X \varphi \, d\mu_i\right) \right| \\ &\leq \sum_{i \in I} \mu_i(X) g\left(\frac{1}{\mu_i(X)} \int_X \varphi \, d\mu_i\right) \leq \sum_{i \in I} \int_X g \circ \varphi \, d\mu_i = \int_X g \circ \varphi \, d\mu < \infty. \end{aligned}$$

We can see that the conditions of Proposition 5 hold, and therefore, by applying it, we obtain

$$\begin{aligned} f\left(\int_X \varphi \, d\mu\right) &= f\left(\sum_{i \in I} p_i f(v_i)\right) \leq \sum_{i \in I} \left(\sum_{j \in J} \lambda_j p_{\pi_j(i)}\right) f\left(\frac{\sum_{j \in J} \lambda_j p_{\pi_j(i)} v_{\pi_j(i)}}{\sum_{j \in J} \lambda_j p_{\pi_j(i)}}\right) \\ &= \sum_{i \in I} \left(\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}(X)\right) f\left(\frac{\sum_{j \in J} \lambda_j \int_X \varphi \, d\mu_{\pi_j(i)}}{\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}(X)}\right) \leq \sum_{i \in I} p_i f(v_i) \\ &= \sum_{i \in I} \mu_i(X) f\left(\frac{1}{\mu_i(X)} \int_X \varphi \, d\mu_i\right). \end{aligned} \tag{5}$$

As a final step, we can apply the integral Jensen’s inequality in (5).

The proof is complete. □

A useful consequence of the previous theorem is the next result.

**Corollary 9** Assume  $(H_1-H_3)$  and  $(H_5)$ . Let  $C \subset \mathbb{R}$  be an interval and  $f : C \rightarrow \mathbb{R}$  be a convex function. Let  $\varphi$  be a  $\mu$ -integrable function on  $X$  taking values in  $C$  such that  $f \circ \varphi$  is also  $\mu$ -integrable on  $X$ . Then

$$\begin{aligned}
 f\left(\int_X \varphi d\mu\right) &\leq C_{set} = C_{set}(f, \varphi, \lambda, \pi, \mathfrak{S}_I) \\
 &:= \sum_{i \in I} \left(\sum_{j \in J} \lambda_j \mu(A_{\pi_j(i)})\right) f\left(\frac{\sum_{j \in J} \lambda_j \int_{A_{\pi_j(i)}} \varphi d\mu}{\sum_{j \in J} \lambda_j \mu(A_{\pi_j(i)})}\right) \leq \int_X f \circ \varphi d\mu. \tag{6}
 \end{aligned}$$

*Proof* Let the measure  $\mu_i$  ( $i \in I$ ) be defined on  $\mathcal{A}$  by

$$\mu_i(A) := \mu(A \cap A_i), \quad A \in \mathcal{A},$$

and then apply Theorem 8.

The proof is complete. □

#### 4 Discussion

First we study the relationship between Theorem 8 and Proposition 5.

Assume  $(H_2)$  and  $(H_3)$ , and let  $(p_i)_{i \in I}$  represent a positive probability distribution. Define the measure  $\mu$  on the power set  $P(I)$  of  $I$  by

$$\mu := \sum_{i \in I} p_i \varepsilon_i,$$

where  $\varepsilon_i$  ( $i \in I$ ) is the unit mass at  $i$  on  $P(I)$ , and use the measure space  $(I, P(I), \mu)$  in  $(H_1)$ . Let  $C \subset \mathbb{R}$  be an interval,  $f : C \rightarrow \mathbb{R}$  be a convex function, and  $(v_i)_{i \in I}$  be a sequence in  $\mathbb{R}$  such that the series  $\sum_{i \in I} p_i v_i$  and  $\sum_{i \in I} p_i f(v_i)$  are absolutely convergent. Define the function  $\varphi$  on  $I$  by

$$\varphi(i) := v_i.$$

Finally, choose  $A_i := \{i\}$  ( $i \in I$ ).

It is easy to check that under these conditions Corollary 9 is equivalent to Proposition 5.

Now we compare our main result with some recent refinement of the integral Jensen’s inequality.

Let  $(X, \mathcal{A}, \nu)$  be a measure space with  $\nu(X) \in ]0, \infty]$ . For the  $\nu$ -integrable positive  $\nu$ -a.e. weight  $w$ , consider the Lebesgue space

$$L_w(X, \nu) := \left\{ \varphi : X \rightarrow \mathbb{R}, \varphi \text{ is } \nu\text{-measurable and } \int_X |\varphi| w d\nu < \infty \right\}.$$

For the  $\nu$ -integrable positive  $\nu$ -a.e. weight  $w$  and given  $n \geq 2$ , we consider the set  $\mathfrak{B}_k(w)$  of all possible  $n$ -tuples of  $\nu$ -integrable positive  $\nu$ -a.e. weights  $\bar{w} = (w_1, \dots, w_n)$  with the property that  $\sum_{i=1}^n w_i = w$ .

The next result can be found in [5].

**Theorem 10** (see [5], Theorem 2.1) *Let  $f : [m, M] \rightarrow \mathbb{R}$  be a convex function,  $\varphi : X \rightarrow [m, M]$  be a  $\nu$ -measurable function such that  $\varphi, f \circ \varphi \in L_w(X, \nu)$ . Then, for any  $\bar{w} \in \mathfrak{B}_k(w)$ , we have*

$$f\left(\frac{\int_X \varphi w d\nu}{\int_X w d\nu}\right) \leq \frac{1}{\int_X w d\nu} \sum_{i=1}^n f\left(\frac{\int_X \varphi w_i d\nu}{\int_X w_i d\nu}\right) \int_X w_i d\nu \leq \frac{\int_X (f \circ \varphi) w d\nu}{\int_X w d\nu},$$

where  $n \geq 2$ .

*Remark 11* Let the measure  $\mu_i$  ( $i \in I$ ) and  $\mu$  be defined on  $\mathcal{A}$  by

$$\mu_i(A) := \frac{1}{\int_X w d\nu} \int_A w_i d\nu, \quad A \in \mathcal{A},$$

and by

$$\mu(A) := \frac{1}{\int_X w d\nu} \int_A w d\nu, \quad A \in \mathcal{A}. \tag{7}$$

Then  $\mu_i(X) > 0$  ( $i \in I$ ) and  $\mu = \sum_{i=1}^n \mu_i$ . By choosing  $k = 1$  (thus  $\lambda_1 = 1$ ), we can see that Theorem 10 is a simple consequence of Theorem 8.

We say that the family of measurable sets  $F_n(X) = \{A_i\}_{i=1, \dots, n}$  is an  $n$ -division for  $X$  if  $X = \bigcup_{i=1}^n A_i, A_i \cap A_j = \emptyset$  for any  $i, j \in \{1, \dots, n\}$  with  $i \neq j$  and  $\nu(A_i) > 0$  for any  $i \in \{1, \dots, n\}$ . For given  $n \geq 2$ , we denote by  $\mathfrak{D}_n(X)$  the set of all  $n$ -divisions of  $X$ .

The following result appears in [6].

**Theorem 12** (see [6], Theorem 2.1) *Let  $f : [m, M] \rightarrow \mathbb{R}$  be a convex function,  $\varphi : X \rightarrow [m, M]$  be a  $\nu$ -measurable function such that  $\varphi, f \circ \varphi \in L_w(X, \nu)$ . Then, for any  $F_n(X) \in \mathfrak{D}_n(X)$ , we have*

$$f\left(\frac{\int_X f w d\nu}{\int_X w d\nu}\right) \leq \frac{1}{\int_X w d\nu} \sum_{i=1}^n f\left(\frac{\int_{A_i} f w d\nu}{\int_{A_i} w d\nu}\right) \int_{A_i} w d\nu \leq \frac{\int_X (f \circ \varphi) w d\nu}{\int_X w d\nu},$$

where  $n \geq 2$ .

*Remark 13* (a) Define the measure  $\mu$  on  $\mathcal{A}$  by (7). By choosing  $k = 1$  (thus  $\lambda_1 = 1$ ), we can see that Theorem 12 is a simple consequence of Corollary 9. Moreover, it follows that Theorem 12 is contained in Theorem 10.

(b) The main result Theorem 2.1 in [7] is the special case of Corollary 9 when  $n = 2$  and  $k = 1$ .

### 5 Applications

Let  $[a, b] \subset \mathbb{R}$  ( $a < b$ ). The  $\sigma$ -algebra of Lebesgue-measurable subsets of  $\mathbb{R}$  is denoted by  $\mathcal{L}$ .  $\lambda$  means the Lebesgue measure on  $\mathcal{L}$ . Assume that  $f : [a, b] \rightarrow \mathbb{R}$  is a convex function and  $g : [a, b] \rightarrow [0, \infty[$  is a Lebesgue-integrable function which is symmetric to  $\frac{a+b}{2}$ . The classical Fejér inequality (see [8]) says

$$f\left(\frac{a+b}{2}\right) \int_a^b g \leq \int_a^b fg \leq \frac{f(a)+f(b)}{2} \int_a^b g. \tag{8}$$

This is a weighted generalization of the Hermite–Hadamard inequality (see [9]) which has the form

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a} \int_a^b f \leq \frac{f(a)+f(b)}{2}. \tag{9}$$

By applying our main result, we can obtain a refinement of the left-hand side of the Fejér inequality. There are refinements of the Fejér inequality (see e.g. [21]), but the next result provides a totally different refinement.

**Proposition 14** *Assume (H<sub>2</sub>–H<sub>3</sub>). Let  $[a, b] \subset \mathbb{R}$  ( $a < b$ ), and consider the measure space  $([a, b], \mathcal{L}, \lambda)$  in (H<sub>1</sub>). Let  $f : [a, b] \rightarrow \mathbb{R}$  be a convex function and  $g : [a, b] \rightarrow [0, \infty[$  be a Lebesgue-integrable function which is symmetric to  $\frac{a+b}{2}$ .*

(a) *If  $\mu := g\lambda$ , where  $g\lambda$  is the measure on  $\mathcal{L}$  having density  $g$  with respect to  $\lambda$ , and  $\mu_i$  ( $i \in I$ ) are measures on  $\mathcal{L}$  such that  $\mu_i([a, b]) > 0$  for all  $i \in I$  and  $\sum_{i \in I} \mu_i = \mu$ , then*

$$f\left(\frac{a+b}{2}\right) \int_a^b g \leq \sum_{i \in I} \left( \sum_{j \in J} \lambda_j \mu_{\pi_j(i)}([a, b]) \right) \cdot f\left(\frac{\sum_{j \in J} \lambda_j \int_{[a, b]} t d\mu_{\pi_j(i)}(t)}{\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}([a, b])}\right) \leq \int_a^b fg.$$

(b) *If  $g_i : [a, b] \rightarrow [0, \infty[$  ( $i \in I$ ) is a Lebesgue-integrable function which is symmetric to  $\frac{a+b}{2}$ , and  $\sum_{i \in I} g_i = g$ , then*

$$f\left(\frac{a+b}{2}\right) \int_a^b g \leq f\left(\frac{a+b}{2}\right) \sum_{i \in I} \left( \sum_{j \in J} \lambda_j \int_a^b g_{\pi_j(i)} \right) \leq \int_a^b fg.$$

*Proof* (a) The result follows immediately from Theorem 8 by choosing  $\varphi : [a, b] \rightarrow \mathbb{R}$ ,  $\varphi(t) = t$ .

(b) It comes from (a) by some easy calculations.

The proof is complete. □

*Remark 15* It is worth to mention that in part (a) the measures may not be symmetric to  $\frac{a+b}{2}$ .

By applying the previous result, we obtain refinements of the Hermite–Hadamard inequality.

**Proposition 16** *Assume (H<sub>2</sub>–H<sub>3</sub>). Let  $C := [a, b] \subset \mathbb{R}$  ( $a < b$ ), and consider the measure space  $([a, b], \mathcal{L}, \lambda)$  in (H<sub>1</sub>). Let  $f : [a, b] \rightarrow \mathbb{R}$  be a convex function.*

(a) *If  $\mu_i$  ( $i \in I$ ) is a measure on  $\mathcal{L}$  such that  $\mu_i([a, b]) > 0$  for all  $i \in I$  and  $\sum_{i \in I} \mu_i = \lambda$ , then*

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a} \sum_{i \in I} \left( \sum_{j \in J} \lambda_j \mu_{\pi_j(i)}([a, b]) \right) \cdot f\left(\frac{\sum_{j \in J} \lambda_j \int_{[a, b]} t d\mu_{\pi_j(i)}(t)}{\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}([a, b])}\right) \leq \frac{1}{b-a} \int_a^b f.$$



(b) Let  $(A_i)_{i \in I}$  be a sequence of pairwise disjoint sets  $A_i \in \mathcal{A}$  with  $\lambda(A_i) > 0$  for all  $i \in I$  and  $\bigcup_{i \in I} A_i = [a, b]$ . Then

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a} \sum_{i \in I} \left( \sum_{j \in J} \lambda_j \lambda(A_{\pi_j(i)}) \right) \cdot f\left(\frac{\sum_{j \in J} \lambda_j \int_{A_{\pi_j(i)}} t dt}{\sum_{j \in J} \lambda_j \lambda(A_{\pi_j(i)})}\right) \leq \frac{1}{b-a} \int_a^b f.$$

(c) If  $a = x_0 < x_1 < \dots < x_n = b$  is a partition of  $[a, b]$ , then

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a} \sum_{i \in I} \left( \sum_{j \in J} \lambda_j (x_{\pi_j(i)} - x_{\pi_j(i)-1}) \right) \cdot f\left(\frac{1}{2} \frac{\sum_{j \in J} \lambda_j (x_{\pi_j(i)}^2 - x_{\pi_j(i)-1}^2)}{\sum_{j \in J} \lambda_j (x_{\pi_j(i)} - x_{\pi_j(i)-1})}\right) \leq \frac{1}{b-a} \int_a^b f.$$

*Proof* (a) This is a special case of Proposition 14 (a).

(b) Let the measure  $\mu_i$  ( $i \in I$ ) be defined on  $\mathcal{L}$  by

$$\mu_i(A) := \lambda(A \cap A_i), \quad A \in \mathcal{L},$$

and then apply (a).

(c) Let  $A_i := [x_{i-1}, x_i[$  ( $i = 1, \dots, n - 1$ ) and  $A_n := [x_{n-1}, x_n]$  in (b).

The proof is complete. □

The second application concerns quasi-arithmetic means.

Let  $C \subset \mathbb{R}$  be an interval, and let  $q : C \rightarrow \mathbb{R}$  be a continuous and strictly monotone function. If  $(X, \mathcal{A}, \mu)$  is a probability space, and  $\varphi : X \rightarrow C$  is a function such that  $q \circ \varphi$  is  $\mu$ -integrable on  $X$ , then

$$M_q(\varphi, \mu) := q^{-1}\left(\int_X q \circ \varphi d\mu\right)$$

is called the quasi-arithmetic mean (integral  $q$ -mean) of  $\varphi$ .

Now we introduce some new quasi-arithmetic means related to the formula  $C_{mes}$ .

**Definition 17** Assume  $(H_1-H_4)$ . Let  $C \subset \mathbb{R}$  be an interval, let  $q, r : C \rightarrow \mathbb{R}$  be continuous and strictly monotone functions, and  $\varphi : X \rightarrow C$  be a  $\mu$ -integrable function for which  $q \circ \varphi$  and  $r \circ \varphi$  are also  $\mu$ -integrable functions. Then we define the following quasi-arithmetic mean of  $\varphi$  with respect to  $C_{mes}$ :

$$M_{q,r}(\varphi, \mu, \lambda, \pi, \mathfrak{M}_I) := q^{-1}\left(\sum_{i \in I} \left(\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}(X)\right) q \circ r^{-1}\left(\frac{\sum_{j \in J} \lambda_j \int_X r \circ \varphi d\mu_{\pi_j(i)}}{\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}(X)}\right)\right).$$

In the next result the introduced means are compared.

**Proposition 18** *Assume that the conditions in Definition 17 hold. If either  $q \circ r^{-1}$  is convex and  $q$  is strictly increasing, or  $q \circ r^{-1}$  is concave and  $q$  is strictly decreasing, then*

$$M_r(\varphi, \mu) \leq M_{q,r}(\varphi, \mu, \lambda, \pi, \mathfrak{M}_I) \leq M_q(\varphi, \mu),$$

*while if either  $r \circ q^{-1}$  is convex and  $r$  is strictly decreasing, or  $r \circ q^{-1}$  is concave and  $r$  is strictly increasing, then*

$$M_q(\varphi, \mu) \leq M_{r,q}(\varphi, \mu, \lambda, \pi, \mathfrak{M}_I) \leq M_r(\varphi, \mu).$$

*Proof* We consider only the case when  $q \circ r^{-1}$  is convex and  $q$  is strictly increasing. By applying Theorem 8 with  $f := q \circ r^{-1}$  and with  $r \circ \varphi$  instead of  $\varphi$ , we obtain

$$\begin{aligned} & q \circ r^{-1} \left( \int_X r \circ \varphi \, d\mu \right) \\ & \leq \sum_{i \in I} \left( \sum_{j \in J} \lambda_j \mu_{\pi_j(i)}(X) \right) q \circ r^{-1} \left( \frac{\sum_{j \in J} \lambda_j \int_X r \circ \varphi \, d\mu_{\pi_j(i)}}{\sum_{j \in J} \lambda_j \mu_{\pi_j(i)}(X)} \right) \leq \int_X q \circ \varphi \, d\mu, \end{aligned}$$

and this implies the result since  $q^{-1}$  is strictly increasing.

The proof is complete. □

Finally, some applications to information theory are presented.

Throughout the rest of the paper probability measures  $P$  and  $Q$  are defined on a fixed measurable space  $(X, \mathcal{A})$ . It is also assumed that  $P$  and  $Q$  are absolutely continuous with respect to a  $\sigma$ -finite measure  $\nu$  on  $\mathcal{A}$ . The densities (or Radon–Nikodym derivatives) of  $P$  and  $Q$  with respect to  $\nu$  are denoted by  $p$  and  $q$ , respectively. These densities are  $\nu$ -almost everywhere uniquely determined.

Introduce the set of functions

$$F := \{f : ]0, \infty[ \rightarrow \mathbb{R} \mid f \text{ is convex}\},$$

and define, for every  $f \in F$ , the function

$$f^* : ]0, \infty[ \rightarrow \mathbb{R}, \quad f^*(t) := tf \left( \frac{1}{t} \right).$$

If  $f \in F$ , then either  $f$  is monotonic or there exists a point  $t_0 \in ]0, \infty[$  such that  $f$  is decreasing on  $]0, t_0[$ . This implies that the limit

$$\lim_{t \rightarrow 0^+} f(t)$$

exists in  $]-\infty, \infty]$ , and

$$f(0) := \lim_{t \rightarrow 0^+} f(t)$$

extends  $f$  into a convex function on  $[0, \infty[$ .

It is well known that, for every  $f \in F$ , the function  $f^*$  also belongs to  $F$ , and therefore

$$f^{**}(0) := \lim_{t \rightarrow 0^+} f^*(t) = \lim_{u \rightarrow \infty} \frac{f(u)}{u}.$$

The important notions of  $f$ -divergence were introduced in [3], [4], and independently in [1].

**Definition 19** (a) For every  $f \in F$ , we define the  $f$ -divergence of  $P$  and  $Q$  by

$$D_f(P, Q) := \int_X q(\omega) f\left(\frac{p(\omega)}{q(\omega)}\right) d\nu(\omega),$$

where the following conventions are used:

$$0f\left(\frac{x}{0}\right) := xf^*(0) \quad \text{if } x > 0, \quad 0f\left(\frac{0}{0}\right) = 0f^*(0) := 0. \tag{10}$$

(b) Let  $f \in F$  be a positive convex function, and let  $\mathbf{p} := (p_1, \dots, p_n)$  and  $\mathbf{q} := (q_1, \dots, q_n)$  represent positive discrete probability distributions. The  $f$ -divergence functional of  $\mathbf{p}$  and  $\mathbf{q}$  is

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

It is possible to use nonnegative discrete probability distributions in the  $f$ -divergence functional by defining

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

*Remark 20* (a) For every  $f \in F$ , the perspective  $\hat{f} : ]0, \infty[ \times ]0, \infty[ \rightarrow \mathbb{R}$  of  $f$  is defined by

$$\hat{f}(x, y) := yf\left(\frac{x}{y}\right).$$

Then (see [18])  $\hat{f}$  is also a convex function. Vajda [22] proved that (10) is the unique rule leading to convex and lower semicontinuous extension of  $\hat{f}$  to the set

$$\{(x, y) \in \mathbb{R}^2 \mid x, y \geq 0\}.$$

(b) Since  $f^*(0) \in ]-\infty, \infty]$ , Lemma 2.8 in [14] shows that  $D_f(P, Q)$  exists in  $]-\infty, \infty]$  and

$$D_f(P, Q) = \int_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega) + f^*(0)P(q = 0). \tag{11}$$

It follows that if  $P$  is absolutely continuous with respect to  $Q$ , then

$$D_f(P, Q) = \int_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega).$$

The basic inequality (see [15])

$$D_f(P, Q) \geq f(1) \tag{12}$$

is one of the key properties of  $f$ -divergences. In the next result a refinement of this inequality is obtained.

**Proposition 21** *Assume (H<sub>2</sub>) and (H<sub>3</sub>). Suppose that we are given a sequence  $\mathfrak{M}_I = (Q_i)_{i \in I}$  of measures on  $\mathcal{A}$  with  $Q_i(X) > 0$  for all  $i \in I$  and  $\sum_{i \in I} Q_i = Q$ . Then*

$$D_f(P, Q) \geq \sum_{i \in I} \left( \sum_{j \in J} \lambda_j Q_{\pi_j(i)}(q > 0) \right) f \left( \frac{\sum_{j \in J} \lambda_j \int_{(q>0)} \frac{p}{q} dQ_{\pi_j(i)}}{\sum_{j \in J} \lambda_j Q_{\pi_j(i)}(q > 0)} \right) + f^*(0)P(q = 0) \geq f \left( \int_{(q>0)} p dv \right) + f^*(0)P(q = 0) \geq f(1). \tag{13}$$

*Proof* Since  $Q$  is absolutely continuous with respect to  $\nu$ , it follows from  $\sum_{i \in I} Q_i = Q$  that  $Q_i$  is also absolutely continuous with respect to  $\nu$  for all  $i \in I$ . By denoting the density (or Radon–Nikodym derivative) of  $Q_i$  with respect to  $\nu$  by  $q_i$  ( $i \in I$ ), we also obtain that  $\sum_{i \in I} q_i = q$   $\nu$ -a.e.

If  $D_f(P, Q) = \infty$ , then (13) is obvious.

If  $D_f(P, Q) \in \mathbb{R}$ , then the integral

$$\int_{(q>0)} f \left( \frac{p(\omega)}{q(\omega)} \right) dQ(\omega) \tag{14}$$

is finite, and therefore either  $Q(p = 0) = 0$  or  $Q(p = 0) > 0$  and  $f(0)$  is finite. It can be seen that the integral Jensen’s inequality can be applied to this integral, and we have from Theorem 8 that

$$D_f(P, Q) \geq \sum_{i \in I} \left( \sum_{j \in J} \lambda_j Q_{\pi_j(i)}(q > 0) \right) f \left( \frac{\sum_{j \in J} \lambda_j \int_{(q>0)} \frac{p}{q} dQ_{\pi_j(i)}}{\sum_{j \in J} \lambda_j Q_{\pi_j(i)}(q > 0)} \right) + f^*(0)P(q = 0) \geq f \left( \int_{(q>0)} p dv \right) + f^*(0)P(q = 0). \tag{15}$$

It follows from the proof of Theorem 2.12 (a) in [14] that

$$f \left( \int_{(q>0)} p dv \right) + f^*(0)P(q = 0) \geq f(1).$$

The proof is complete. □

*Remark 22* (a) By using density functions, (13) can be rewritten in the following form:

$$D_f(P, Q) \geq \sum_{i \in I} \left( \sum_{j \in J} \lambda_j \int_{(q>0)} q_{\pi_j(i)} dv \right) f \left( \frac{\sum_{j \in J} \lambda_j \int_{(q>0)} \frac{p}{q} q_{\pi_j(i)} dv}{\sum_{j \in J} \lambda_j \int_{(q>0)} q_{\pi_j(i)} dv} \right) + f^*(0)P(q = 0) \geq f \left( \int_{(q>0)} p dv \right) + f^*(0)P(q = 0) \geq f(1).$$

(b) Define

$$q_i := \sum_{j \in J} \lambda_j Q_{\pi_j(i)}(q > 0), \quad i \in I, \quad \widehat{q} := Q(q = 0),$$

and

$$p_i := \sum_{j \in J} \lambda_j \int_{(q>0)} \frac{p}{q} dQ_{\pi_j(i)}, \quad i \in I, \quad \widehat{p} := P(q = 0).$$

By applying (H<sub>3</sub>), (H<sub>4</sub>), Theorem 7, and  $\sum_{i \in I} Q_i = Q$ , we have

$$\begin{aligned} \sum_{i \in I} q_i &= \sum_{i \in I} \left( \sum_{j \in J} \lambda_j Q_{\pi_j(i)}(q > 0) \right) = \sum_{j \in J} \lambda_j \left( \sum_{i \in I} Q_{\pi_j(i)}(q > 0) \right) = \\ &= \sum_{j \in J} \lambda_j \left( \sum_{i \in I} Q_i(q > 0) \right) = Q(q > 0), \end{aligned}$$

and hence  $(q_i)_{i \in I}$  and  $\widehat{q}$  represent a discrete probability distribution  $\mathbf{q}$ .

Similarly, it follows from (H<sub>3</sub>), (H<sub>4</sub>), Theorem 7, and  $\sum_{i \in I} Q_i = Q$ , that

$$\begin{aligned} \sum_{i \in I} p_i &= \sum_{i \in I} \left( \sum_{j \in J} \lambda_j \int_{(q>0)} \frac{p}{q} dQ_{\pi_j(i)} \right) = \sum_{j \in J} \lambda_j \left( \sum_{i \in I} \int_{(q>0)} \frac{p}{q} dQ_{\pi_j(i)} \right) \\ &= \sum_{j \in J} \lambda_j \left( \sum_{i \in I} \int_{(q>0)} \frac{p}{q} dQ_i \right) = \int_{(q>0)} \frac{p}{q} dQ = P(q > 0), \end{aligned}$$

and therefore  $(p_i)_{i \in I}$  and  $\widehat{p}$  also represent a discrete probability distribution  $\mathbf{p}$ . Since

$$0f\left(\frac{\widehat{p}}{\widehat{q}}\right) := \begin{cases} 0 = f^*(0)P(q = 0), & \widehat{p} = 0, \\ \lim_{t \rightarrow 0^+} tf\left(\frac{\widehat{p}}{t}\right) = f^*(0)P(q = 0), & \widehat{p} > 0, \end{cases}$$

we can see that the second term in (13) can be considered as the  $f$ -divergence functional of  $\mathbf{p}$  and  $\mathbf{q}$  (if  $I = \mathbb{N}_+$ , then we can say that  $I_f(\mathbf{p}, \mathbf{q})$  is a generalized  $f$ -divergence functional). It is interesting that inequality (12) for “continuous”  $f$ -divergence can be refined by “discrete”  $f$ -divergence.

**Acknowledgements**

Not applicable.

**Funding**

The research of the author was supported by Hungarian Scientific Research Fund Grant No. K101217 and Széchenyi 2020 under EFOP-3.6.1-16-2016-00015.

**Availability of data and materials**

Not applicable.

**Competing interests**

The author declares that they have no competing interests.

**Authors’ contributions**

The author carried out the results and read and approved the current version of the manuscript.

## Publisher's Note

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

Received: 30 June 2020 Accepted: 5 January 2021 Published online: 14 January 2021

## References

1. Ali, M.S., Silvey, D.: A general class of coefficients of divergence of one distribution from another. *J. R. Stat. Soc. Ser. B* **28**, 131–140 (1966)
2. Brnetić, I., Khan, K.A., Pečarić, J.: Refinement of Jensen's inequality with applications to cyclic mixed symmetric means and Cauchy means. *J. Math. Inequal.* **9**(4), 1309–1321 (2015)
3. Csiszár, I.: Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität on Markoffschen Ketten. *Publ. Math. Inst. Hung. Acad. Sci. Ser. A* **8**, 84–108 (1963)
4. Csiszár, I.: Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hung.* **2**, 299–318 (1967)
5. Dragomir, S.S.: Refining Jensen's integral inequality for partitions of weights. *Libertas Math.* **37**, 25–44 (2017)
6. Dragomir, S.S.: Refining Jensen's integral inequality for divisions of measurable space. *J. Math. Ext.* **12**(2), 87–106 (2018)
7. Dragomir, S.S., Khan, M.A., Abathun, A.: Refinement of the Jensen integral inequality. *Open Math.* **14**, 221–228 (2016)
8. Fejér, L.: Über die Fourierreihen, II. *Math. Naturwiss. Anz. Ungar. Akad. Wiss.* **24**, 369–390 (1906) (in Hungarian)
9. Hadamard, J.: Étude sur les propriétés des fonctions entières en particulier d'une fonction considérée par Riemann. *J. Math. Pures Appl.* **58**, 171–215 (1893)
10. Horváth, L.: A refinement of the integral form of Jensen's inequality. *J. Inequal. Appl.* **2012**, Article ID 178 (2012)
11. Horváth, L.: New refinements of the discrete Jensen's inequality generated by finite or infinite permutations. *Aequ. Math.* <https://doi.org/10.1007/s00010-019-00696-z>
12. Horváth, L., Khan, K.A., Pečarić, J.: Combinatorial Improvements of Jensen's Inequality. *Element*, Zagreb (2014)
13. Horváth, L., Khan, K.A., Pečarić, J.: Cyclic refinements of the discrete and integral form of Jensen's inequality with applications. *Analysis* **36**(4), 253–262 (2016)
14. Horváth, L., Pečarić, D., Pečarić, J.: A refinement and an exact equality condition for the basic inequality of  $f$ -divergences. *Filomat* **32**(12), 4263–4273 (2018)
15. Liese, F., Vajda, I.: On divergences and informations in statistics and information theory. *IEEE Trans. Inf. Theory* **52**, 4394–4412 (2006)
16. Niculescu, C., Persson, L.E.: *Convex Functions and Their Applications. A Contemporary Approach*. Springer, Berlin (2006)
17. Perlman, M.D.: Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *J. Multivar. Anal.* **4**, 52–65 (1974)
18. Rockafellar, T.: *Convex Analysis*. Princeton University Press, Princeton (1970)
19. Rooin, J.: A refinement of Jensen's inequality. *J. Inequal. Pure Appl. Math.* **6**(2), Article ID 38 (2005)
20. Rudin, W.: *Real and Complex Analysis*, 3rd edn. McGraw-Hill, New York (1987)
21. Tseng, K.-L., Hwang, S.-R., Dragomir, S.S.: Refinements of Fejér's inequality for convex functions. *Period. Math. Hung.* **65**, 17–28 (2012)
22. Vajda, I.: *Theory of Statistical Inference and Information*. Kluwer Academic, Boston (1989)

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

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

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---