Filomat 32:12 (2018), 4263–4273
https://doi.org/10.2298/FIL1812263H

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

A Refinement and an Exact Equality Condition for the Basic Inequality of f-divergences

László Horváth[a], Đilda Pečarić[b], Josip Pečarić[c]

[a]*Department of Mathematics, University of Pannonia*
[b]*Catholic University of Croatia*
[c]*Faculty of Textile Technology, University of Zagreb*

Abstract. f-divergences play important role in probability theory, especially in information theory and in mathematical statistics. Remarkable divergences can be found among them. Inequalities for f-divergences are very useful and applicable in information theory. In this paper we give a precise equality condition and a refinement for one of the basic inequalities of f-divergences. The results are illustrated by some applications.

1. Introduction

Measures of dissimilarity between probability measures play important role in probability theory, especially in information theory and in mathematical statistics. Many divergence measures for this purpose have been introduced and studied (see for example Vajda [14]). Among them f-divergences (see Section 2 for exact definitions) were introduced by Csiszár [2]-[3] and independently by Ali and Silvey [1]. Remarkable divergences can be found among f-divergences, such as the information divergence, the Pearson or χ^2-divergence, the Hellinger distance and total variational distance. There are a lot of papers dealing with f-divergence inequalities (see Dragomir [5], Dembo, Cover, and Thomas [4] and Sason and Verdú [13]). These inequalities are very useful and applicable in information theory.

One of the basic inequalities is (see Liese and Vajda [10])

$$D_f(P,Q) \geq f(1).$$

In this paper we give a refinement and a precise equality condition for this inequality. Some applications for discrete distributions, for the Shannon entropy, and some examples are given.

2. Preliminary Results and Basic Definitions

The classical Jensen's inequality is well known (see [7]).

2010 *Mathematics Subject Classification*. 26D15; 94A17
Keywords. f-divergence, inequalities for f-divergence, Jensen's inequality, Shannon entropy
Received: 27 November 2017; Accepted: 17 June 2018
Communicated by Fuad Kittaneh
The research of the first author has been supported by Hungarian National Foundations for Scientific Research Grant No. K120186.
Email addresses: lhorvath@almos.uni-pannon.hu (László Horváth), gildapeca@gmail.com (Đilda Pečarić), pecaric@mahazu.hazu.hr (Josip Pečarić)

Theorem 2.1. *Let g be an integrable function on a probability space $(Y, \mathcal{B}, \nu)$ taking values in an interval $I \subset \mathbb{R}$. Then $\int_Y g d\nu$ lies in I. If f is a convex function on I such that $f \circ g$ is ν-integrable, then*

$$f\left(\int_Y g d\nu\right) \leq \int_Y f \circ g d\nu. \tag{1}$$

The following approach to give a necessary and sufficient condition for equality in this inequality may be new. First, we introduce the next definition.

Definition 2.2. *Let $(Y, \mathcal{B}, \nu)$ be a probability space, and let g be a real measurable function defined almost everywhere on Y. We denote by $essint_\nu(g)$ the smallest interval in $\mathbb{R}$ for which*

$$\nu(g \in essint_\nu(g)) = 1.$$

Remark 2.3. *(a) Obviously, the endpoints of $essint_\nu(g)$ are the essential infimum ($essinf_\nu(g)$) and the essential supremum of g, and either of them belong to $essint_\nu(g)$ exactly if g takes this value with positive probability.*

(b) It is easy to see that either $essint_\nu(g) = \left\{\int_Y g d\nu\right\}$ (in this case g is constant ν-a.e.) or $\int_Y g d\nu$ is an inner point of $essint_\nu(g)$.

(c) The interval $essinf_\nu(g)$ is connected with the essential range of g, but not the same set (for example, the essential range of g is always closed, and not an interval in general).

Lemma 2.4. *Assume the conditions of Theorem 2.1 are satisfied. Equality holds in (1) if and only if f is affine on $essint_\nu(g)$.*

Proof. It is easy to see that the condition is sufficient for equality in (1).

Conversely, if $\text{essint}_\nu(g)$ contains only one point, then it is trivial, so we can assume that $m := \int_Y g d\nu$ is an inner point of $\text{essint}_\nu(g)$. Let

$$l : \mathbb{R} \to \mathbb{R}, \quad l(t) = f'_+(m)(t - m) + f(m).$$

If f is not affine on $\text{essint}_\nu(g)$, then by the convexity of f, there is a point $t_1 \in \text{essint}_\nu(g)$ such that $f(t_1) > l(t_1)$. Suppose $t_1 > m$ (the case $t_1 < m$ can be handled similarly). Since f is convex, $f(t) \geq l(t)$ $(t \in I)$ and $f(t) > l(t)$ $(t \in I,\ t \geq t_1)$. It follows by using $\nu(g > t_1) > 0$, that

$$\int_Y f \circ g d\nu = \int_{(g<t_1)} f \circ g d\nu + \int_{(g\geq t_1)} f \circ g d\nu$$

$$\geq \int_{(g<t_1)} l \circ g d\nu + \int_{(g\geq t_1)} f \circ g d\nu > \int_Y l \circ g d\nu = f(m),$$

which is a contradiction.

The proof is complete. □

The next refinement of the Jensen's inequality can be found in Horváth [8].

Theorem 2.5. *Let $I \subset \mathbb{R}$ be an interval, and let $f : I \to \mathbb{R}$ be a convex function. Let $(Y, \mathcal{B}, \nu)$ be a probability space, and let $g : Y \to I$ be a ν-integrable function such that $f \circ g$ is also ν-integrable. Suppose that $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n} \alpha_i = 1$. Then*

(a)

$$f\left(\int_Y g d\nu\right) \leq \int_{Y^n} f\left(\sum_{i=1}^{n} \alpha_i g(x_i)\right) d\nu^n(x_1, \ldots, x_n) \leq \int_Y f \circ g d\nu.$$

(b)

$$\int_{Y^{n+1}} f\left(\frac{1}{n+1}\sum_{i=1}^{n+1} g(x_i)\right) d\nu^{n+1}(x_1, \ldots, x_{n+1})$$

$$\leq \int_{Y^n} f\left(\frac{1}{n}\sum_{i=1}^{n} g(x_i)\right) d\nu^n(x_1, \ldots, x_n) \leq \int_{Y^n} f\left(\sum_{i=1}^{n} \alpha_i g(x_i)\right) d\nu^n(x_1, \ldots, x_n).$$

By analyzing the proof of the previous result, it can be seen that the hypothesis "$f \circ g$ is ν-integrable" can be weaken.

Theorem 2.6. *Let $I \subset \mathbb{R}$ be an interval, and let $f : I \to \mathbb{R}$ be a convex function. Let $(Y, \mathcal{B}, \nu)$ be a probability space, and let $g : Y \to I$ be a ν-integrable function such that the integral $\int_Y f \circ g d\nu$ exists in $]-\infty, \infty]$. Suppose that $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n} \alpha_i = 1$. Then the assertions of Theorem 2.5 remain true.*

We assume throughout that the 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 σ-finite measure μ on $\mathcal{A}$. The densities (or Radon-Nikodym derivatives) of P and Q with respect to μ are denoted by p and q, respectively. These densities are μ-almost everywhere uniquely determined.

Let

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

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

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

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

$$\lim_{t \to 0+} f(t)$$

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

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

extends f into a convex function on $[0, \infty[$. The extended function is continuous and has finite left and right derivatives at each point of $]0, \infty[$.

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

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

We need the following simple property of functions belonging to F.

Lemma 2.7. *If $f \in F$, then $f^*(0) \geq f'_+(1)$. This inequality becomes an equality if and only if*

$$f(t) = f'_+(1)(t-1) + f(1), \quad t \geq 1. \tag{2}$$

Proof. Since f is convex,

$$f(t) \geq f'_+(1)(t-1) + f(1), \quad t \geq 1,$$

and therefore

$$f^*(0) = \lim_{t \to \infty} \frac{f(t)}{t} \geq f'_+(1).$$

If (2) is satisfied, then obviously $f^*(0) = f'_+(1)$.
If there exists $t_1 > 1$ such that $f'_+(t_1) > f'_+(1)$, then by the convexity of f,

$$f(t) \geq f'_+(t_1)(t-t_1) + f(t_1), \quad t \geq t_1,$$

and hence $f^*(0) > f'_+(1)$. It follows that $f^*(0) = f'_+(1)$ implies

$$f'_+(t) = f'_+(1), \quad t \geq t_1,$$

and this gives (2) (see [6] 1.6.2 Corollary 2).
The proof is complete. □

The next result prepares the notion of f-divergence of probability measures.

Lemma 2.8. *For every $f \in F$ the integral*

$$\int_{(q>0)} q(\omega) f\left(\frac{p(\omega)}{q(\omega)}\right) d\mu(\omega)$$

exists and it belongs to the interval $]-\infty, \infty]$.

Proof. Since f is convex,

$$f(t) \geq f'_+(1)(t-1) + f(1), \quad t \geq 0.$$

This implies that for all $\omega \in (q > 0)$

$$q(\omega) f\left(\frac{p(\omega)}{q(\omega)}\right) \geq h(\omega) := f'_+(1)(p(\omega) - q(\omega)) + f(1) q(\omega). \tag{3}$$

Elementary considerations show that the function h is μ-integrable over $(q > 0)$, and this gives the result by (3).
The proof is complete. □

Now we introduce the notion of f-divergence.

Definition 2.9. *For every $f \in F$ we define the f-divergence of P and Q by*

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

where the following conventions are used

$$0f\left(\frac{x}{0}\right) := xf^*(0) \text{ if } x > 0, \quad 0f\left(\frac{0}{0}\right) = 0f^*(0) := 0. \tag{4}$$

Remark 2.10. *(a) For every $f \in F$ the perspective $\hat{f} :]0,\infty[\times]0,\infty[\to \mathbb{R}$ of f is defined by*

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

Then (see [12]) $\hat{f}$ is also a convex function. Vajda [14] proved that (4) is the unique rule leading to convex and lower semicontinuous extension of $\hat{f}$ to the set

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

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

$$D_f(P,Q) = \int_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega) + f^*(0) P(q=0). \tag{5}$$

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

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

Various divergences in information theory and statistics are special cases of the f-divergence. We illustrate this by some examples.

(a) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(t) = t\ln(t)$ in (5), the information divergence is obtained

$$I(P,Q) = \int_{(q>0)} p(\omega) \ln\left(\frac{p(\omega)}{q(\omega)}\right) d\mu(\omega) + \infty P(q=0). \tag{6}$$

(b) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(t) = (t-1)^2$ in (5), the Pearson or χ^2-divergence is obtained

$$\chi^2(P,Q) = \int_{(q>0)} \frac{(p(\omega)-q(\omega))^2}{q(\omega)} d\mu(\omega) + \infty P(q=0). \tag{7}$$

(c) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(t) = \left(\sqrt{t}-1\right)^2$ in (5), the Hellinger distance is obtained

$$H^2(P,Q) = \int_X \left(\sqrt{p(\omega)} - \sqrt{q(\omega)}\right)^2 d\mu(\omega). \tag{8}$$

(d) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(t) = |t-1|$ in (5), the total variational distance is obtained

$$V(P,Q) = \int_X \left|p(\omega) - q(\omega)\right| \mu(\omega). \tag{9}$$

We need the following lemma.

Lemma 2.11. *Let* $t_0 := P(q > 0)$.
(a) For every $\varepsilon > 0$

$$Q\left(\frac{p}{q} < t_0 + \varepsilon,\ q > 0\right) > 0.$$

(b)

$$essinf_Q\left(\frac{p}{q}\right) \leq t_0$$

Proof. (a) Obviously,

$$Q\left(\frac{p}{q} < t_0 + \varepsilon,\ q > 0\right) = 1 - Q\left(\frac{p}{q} \geq t_0 + \varepsilon,\ q > 0\right).$$

The result follows from this, since

$$Q\left(\frac{p}{q} \geq t_0 + \varepsilon,\ q > 0\right) = \int\limits_X q 1_{\left(\frac{p}{q} \geq t_0 + \varepsilon,\ q > 0\right)} d\mu \leq \int\limits_{(q>0)} \frac{1}{t_0 + \varepsilon} p d\mu$$

$$= \frac{t_0}{t_0 + \varepsilon} < 1.$$

(b) It comes from (a).
The proof is complete. □

The following result contains a key property of f-divergences. We give a simple proof which emphasizes the importance of the convexity of f, and give an exact equality condition.

Theorem 2.12. *(a) For every* $f \in F$

$$D_f(P, Q) \geq f(1). \tag{10}$$

(b) Assume $P(q = 0) = 0$. *Then equality holds in (10) if and only if* f *is affine on* $essint_Q\left(\frac{p}{q}\right)$.
(c) Assume $P(q = 0) > 0$. *Then equality holds in (10) if and only if* f *is affine on* $essint_Q\left(\frac{p}{q}\right) \cup [1, \infty[$.

Proof. (a) If $D_f(P, Q) = \infty$, then (10) is obvious.
If $D_f(P, Q) \in \mathbb{R}$, then the integral

$$\int\limits_{(q>0)} f\left(\frac{p(\omega)}{q(\omega)}\right) dQ(\omega) \tag{11}$$

is finite, and therefore either $Q(p = 0) = 0$ or $Q(p = 0) > 0$ and $f(0)$ is finite. It follows that Jensen's inequality can be applied to this integral, and we have

$$D_f(P, Q) \geq f\left(\int\limits_{(q>0)} p d\mu\right) + f^*(0) P(q = 0) \tag{12}$$

$$= f(P(q > 0)) + f^*(0) P(q = 0). \tag{13}$$

Let $t_0 := P(q > 0)$. By using Lemma 2.7, $t_0 \in [0,1]$, and the convexity of f, it follows from (13) that

$$D_f(P,Q) \geq f(t_0) + f'_+(1)(1-t_0) \tag{14}$$

$$\geq f(1) + f'_+(1)(t_0-1) + f'_+(1)(1-t_0) = f(1). \tag{15}$$

(b) If $D_f(P,Q) = f(1)$, then $D_f(P,Q)$ is finite.

Assume $P(q=0) = 0$. Then by (12) and (13), $D_f(P,Q) = f(1)$ is satisfied if and only if equality holds in the Jensen's inequality. Lemma 2.4 shows that this happens exactly if f is affine on $\operatorname{essint}_Q\left(\frac{p}{q}\right)$.

(c) Assume $P(q=0) > 0$. Then (12), (13), (14) and (15) yield that there must be equality in the Jensen's inequality, $f^*(0) = f'_+(1)$, and

$$f(t_0) = f(1) + f'_+(1)(t_0-1). \tag{16}$$

By Lemma 2.4 and Lemma 2.7, the first two equality conditions are satisfied exactly if f is affine on $\operatorname{essint}_Q\left(\frac{p}{q}\right) \cup [1,\infty[$.

Now assume that f is affine on $\operatorname{essint}_Q\left(\frac{p}{q}\right) \cup [1,\infty[$. In case of $t_0 > 0$, Lemma 2.11 (b) and the continuity of f at t_0 show that (16) also holds. In case of $t_0 = 0$, it is easy to see that $Q\left(\frac{p}{q} = 0\right) = 1$, and hence $0 \in \operatorname{essint}_Q\left(\frac{p}{q}\right)$ which implies (16) too.

The proof is complete. □

Remark 2.13. *(a) Consider the subclass $F_1 \subset F$ such that $f \in F_1$ satisfies $f(1) = 0$. In this case inequality (10) has the usual form*

$$D_f(P,Q) \geq 0.$$

(b) The usual equality condition is the next (see [10]): if f is strictly convex at 1, then $D_f(P,Q) = f(1)$ holds if and only if $P = Q$. Theorem 2.12 (b) and (c) give more precise conditions.

3. Main Results

Suppose that $\alpha_1, \ldots, \alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n} \alpha_i = 1$. Let

$$\mathcal{A}^n := \mathcal{A} \otimes \ldots \otimes \mathcal{A}, \quad \text{with } n \text{ factors},$$

and define the probability measures Q^n and R on $\mathcal{A}^n$ by

$$Q^n := Q \otimes \ldots \otimes Q, \quad \text{with } n \text{ factors},$$

and

$$R_\alpha := \sum_{i=1}^{n} \alpha_i Q \otimes \ldots \otimes Q \otimes \overset{i}{\check{P}} \otimes Q \otimes \ldots \otimes Q.$$

In case of $\alpha_i = \frac{1}{n}$ $(i = 1, \ldots, n)$ the probability measure R_α will be denoted by R_n.

These measures are absolutely continuous with respect to μ^n on $\mathcal{A}^n$. The densities of R and Q^n with respect to μ^n are

$$\bigotimes_{i=1}^{n} q : X^n \to \mathbb{R}, \quad (\omega_1, \ldots, \omega_n) \to \prod_{i=1}^{n} q(\omega_i),$$

and

$$(\omega_1,\ldots,\omega_n)\to\sum_{i=1}^{n}\alpha_i q(\omega_1)\ldots\overset{i}{\check{p}}(\omega_i)\ldots q(\omega_n),\quad (\omega_1,\ldots,\omega_n)\in X^n,$$

respectively.

It is easy to calculate that

$$R_\alpha\left(\bigotimes_{i=1}^{n}q=0\right)=1-R_\alpha\left(\bigotimes_{i=1}^{n}q>0\right)=1-R_\alpha\left((q>0)^n\right)$$

$$=1-\sum_{i=1}^{n}\alpha_i Q(q>0)^{n-1}P(q>0)=1-P(q>0)=P(q=0).$$

It follows that for every $f\in F$

$$D_f(R_\alpha,Q^n)=\int\limits_{(q>0)^n}f\left(\frac{\sum_{i=1}^{n}\alpha_i q(\omega_1)\ldots p(\omega_i)\ldots q(\omega_n)}{\prod_{i=1}^{n}q(\omega_i)}\right)dQ^n(\omega_1,\ldots,\omega_n)$$

$$+f^*(0)R_\alpha\left(\bigotimes_{i=1}^{n}q=0\right)$$

$$=\int\limits_{(q>0)^n}f\left(\sum_{i=1}^{n}\alpha_i\frac{p(\omega_i)}{q(\omega_i)}\right)dQ^n(\omega_1,\ldots,\omega_n)+f^*(0)P(q=0) \tag{17}$$

$$=\int\limits_{(q>0)^n}\prod_{i=1}^{n}q(\omega_i)\,f\left(\sum_{i=1}^{n}\alpha_i\frac{p(\omega_i)}{q(\omega_i)}\right)d\mu^n(\omega_1,\ldots,\omega_n)+f^*(0)P(q=0).$$

By applying Theorem 2.5, we obtain some refinements of the basic inequality 10.

Theorem 3.1. *Suppose that $\alpha_1,\ldots,\alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n}\alpha_i=1$. If $f\in F$, then*

(a)

$$D_f(P,Q)\geq D_f(R_\alpha,Q^n)\geq D_f(R_n,Q^n)\geq f(1). \tag{18}$$

(b)

$$D_f(P,Q)=D_f\left(R_1,Q^1\right)$$

$$\geq\ldots\geq D_f(R_m,Q^m)\geq D_f\left(R_{m+1},Q^{m+1}\right)\geq\ldots\geq f(1),\quad m\geq 1.$$

Proof. (a) The third inequality in (18) comes from Theorem 2.12.

So it remains to prove the first two inequalities in (18). By (5) and (17), it is enough to show that

$$\int\limits_{(q>0)}f\left(\frac{p(\omega)}{q(\omega)}\right)dQ(\omega)\geq\int\limits_{(q>0)^n}f\left(\sum_{i=1}^{n}\alpha_i\frac{p(\omega_i)}{q(\omega_i)}\right)dQ^n(\omega_1,\ldots,\omega_n) \tag{19}$$

$$\geq \int\limits_{(q>0)^n} f\left(\frac{1}{n}\sum_{i=1}^{n}\frac{p(\omega_i)}{q(\omega_i)}\right) dQ^n(\omega_1,\ldots,\omega_n),$$

which is an immediate consequence of Theorem 2.6.

(b) We can proceed similarly as in (a).

The proof is complete. □

By considering the special f-divergences (6-9), we have after each other

(a) the information divergence

$$I(R_\alpha, Q^n) = \infty P(q=0)$$

$$+ \int\limits_{(q>0)^n} \sum_{i=1}^{n}\left(\alpha_i p(\omega_i)\prod_{\substack{j=1\\ j\neq i}}^{n} q(\omega_j)\right)\ln\left(\sum_{i=1}^{n}\alpha_i\frac{p(\omega_i)}{q(\omega_i)}\right) d\mu^n(\omega_1,\ldots,\omega_n),$$

(b) the Pearson divergence

$$\chi^2(R_\alpha, Q^n) =$$

$$= \int\limits_{(q>0)^n} \prod_{i=1}^{n} q(\omega_i)\left(\sum_{i=1}^{n}\alpha_i\frac{p(\omega_i)-q(\omega_i)}{q(\omega_i)}\right)^2 d\mu^n(\omega_1,\ldots,\omega_n) + \infty P(q=0),$$

(c) the Hellinger distance

$$H^2(R_\alpha, Q^n) = \int\limits_{(q>0)^n} \prod_{i=1}^{n} q(\omega_i)\left(\left(\sum_{i=1}^{n}\alpha_i\frac{p(\omega_i)}{q(\omega_i)}\right)^{1/2} - 1\right)^2 d\mu^n(\omega_1,\ldots,\omega_n),$$

(d) the total variational distance

$$V(R_\alpha, Q^n) = \int\limits_{(q>0)^n} \prod_{i=1}^{n} q(\omega_i)\left|\sum_{i=1}^{n}\alpha_i\frac{p(\omega_i)-q(\omega_i)}{q(\omega_i)}\right| d\mu^n(\omega_1,\ldots,\omega_n).$$

Now, we consider the special case, important in many applications, in which P and Q are discrete distributions.

Denote T either the set $\{1,\ldots,k\}$ with a fixed positive integer k, or the set $\{1,2,\ldots\}$. We say that P and Q are derived from the positive probability distributions $p := (p_i)_{i\in T}$ and $q := (q_i)_{i\in T}$, respectively, if $p_i, q_i > 0$ $(i \in T)$, and $\sum_{i\in T} p_i = \sum_{i\in T} q_i = 1$. In this case $X = T$, $\mathcal{A}$ is the power set of T, and μ is the counting measure on $\mathcal{A}$.

Corollary 3.2. *Suppose that $\alpha_1,\ldots,\alpha_n$ are nonnegative numbers with $\sum_{i=1}^{n}\alpha_i = 1$. Suppose also that P and Q are derived from the positive probability distributions $(p_i)_{i\in T}$ and $(q_i)_{i\in T}$, respectively. If $f \in F$, then*

(a)

$$D_f(P,Q) = \sum_{i\in T} q_i f\left(\frac{p_i}{q_i}\right) \geq \sum_{(i_1,\ldots,i_n)\in T^n}\prod_{j=1}^{n} q_{i_j} f\left(\sum_{j=1}^{n}\alpha_j\frac{p_{i_j}}{q_{i_j}}\right)$$

$$\geq \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} f\left(\frac{1}{n}\sum_{j=1}^{n}\frac{p_{i_j}}{q_{i_j}}\right) \geq f(1).$$

(b)

$$D_f(P,Q) \geq \ldots \geq \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} f\left(\frac{1}{n}\sum_{j=1}^{n}\frac{p_{i_j}}{q_{i_j}}\right)$$

$$\geq \sum_{(i_1,\ldots,i_{n+1})\in T^{n+1}} \prod_{j=1}^{n+1} q_{i_j} f\left(\frac{1}{n+1}\sum_{j=1}^{n+1}\frac{p_{i_j}}{q_{i_j}}\right) \geq \ldots \geq f(1), \quad n \geq 1.$$

Proof. This comes from Theorem 3.1 immediately. □

Finally, we give an example to illustrate the previous result. We consider only Corollary 3.2 (a).

Example 3.3. *(a) By choosing $f :]0,\infty[\to \mathbb{R}$, $f(x) = -\ln(x)$ and $p_i = \frac{1}{k}$ $(i = 1,\ldots,k)$ in the previous corollary (in this case $T = \{1,\ldots,k\}$), we have*

$$D_f(P,Q) = -\sum_{i=1}^{k} q_i \ln\left(\frac{1}{kq_i}\right) = \ln(k) + \sum_{i=1}^{k} q_i \ln(q_i)$$

$$\geq -\sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\frac{1}{k}\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right) = \ln(k) - \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right)$$

$$\geq -\sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\frac{1}{kn}\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right)$$

$$= \ln(kn) - \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right) \geq 0.$$

It can be obtained from this some refinements of the classical upper estimation for the Shannon entropy

$$H(Q) := -\sum_{i=1}^{k} q_i \ln(q_i) \leq \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{\alpha_j}{q_{i_j}}\right)$$

$$\leq -\ln(n) + \sum_{(i_1,\ldots,i_n)\in T^n} \prod_{j=1}^{n} q_{i_j} \ln\left(\sum_{j=1}^{n}\frac{1}{q_{i_j}}\right) \leq \ln(k).$$

(b) If $f :]0,\infty[\to \mathbb{R}$, $f(x) = x\ln(x)$ in the previous corollary, then we have the following estimations for the information or Kullback–Leibler divergence:

$$I(P,Q) = \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right) \geq \sum_{(i_1,\ldots,i_n)\in T^n} \left(\sum_{j=1}^{n} \alpha_j p_{i_j} \prod_{\substack{l=1\\ l\neq j}}^{n} q_{i_l}\right) \ln\left(\sum_{j=1}^{n}\alpha_j\frac{p_{i_j}}{q_{i_j}}\right)$$

$$\geq \frac{1}{n}\sum_{(i_1,\ldots,i_n)\in T^n} \left(\sum_{j=1}^{n} p_{i_j} \prod_{\substack{l=1\\ l\neq j}}^{n} q_{i_l}\right) \ln\left(\frac{1}{n}\sum_{j=1}^{n}\frac{p_{i_j}}{q_{i_j}}\right) \geq 0. \tag{20}$$

(c) The Zipf-Mandelbrot law (see Mandelbrot [11] and Zipf [15]) is a discrete probability distribution depends on three parameters $N \in \{1,2,\ldots\}$, $q \in [0,\infty[$ and $s > 0$, and it is defined by

$$f(i;N,q,s) := \frac{1}{(i+q)^s H_{N,q,s}}, \quad i = 1,\ldots,N,$$

where

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

Let P and Q be the Zipf-Mandelbrot law with parameters $N \in \{1,2,\ldots\}$, $q_1, q_2 \in [0,\infty[$ and $s_1, s_2 > 0$, respectively, and let $2 \le k \le N$ be an integer. It follows from the first part of (20) with $T = \{1,\ldots,N\}$ that

$$I(P,Q) = \sum_{i=1}^{N} \frac{1}{(i+q_1)^{s_1} H_{N,q_1,s_1}} \log\left(\frac{(i+q_2)^{s_2} H_{N,q_2,s_2}}{(i+q_1)^{s_1} H_{N,q_1,s_1}}\right)$$

$$\ge \sum_{(i_1,\ldots,i_N)\in T^n} \left(\sum_{j=1}^{n} \alpha_j \frac{1}{\left(i_j+q_1\right)^{s_1} H_{N,q_1,s_1}} \prod_{\substack{l=1 \\ l\ne j}}^{n} \frac{1}{(i_l+q_2)^{s_2} H_{N,q_2,s_2}}\right)$$

$$\times \ln\left(\sum_{j=1}^{n} \alpha_j \frac{\left(i_j+q_2\right)^{s_2} H_{N,q_2,s_2}}{\left(i_j+q_1\right)^{s_1} H_{N,q_1,s_1}}\right) \ge 0.$$

This is another type of refinement for $I(P,Q)$ than it is given in [9].

References

[1] M. S. Ali and D. Silvey, A general class of coefficients of divergence of one distribution from another, J. Roy. Statist. Soc., ser. B **28** (1966), 131–140.
[2] I. Csiszár, Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität on Markoffschen Ketten, Publ. Math. Inst. Hungar. Acad. Sci., ser. A **8** (1963), 84–108.
[3] I. Csiszár, Information-type measures of difference of probability distributions and indirect observations, Studia Sci. Math. Hungar. **2** (1967), 299–318.
[4] A. Dembo, T.M. Cover and J. A. Thomas, Information theoretic inequalities, IEEE Trans. Inf. Theory **37** (1991), 1501-1518.
[5] S. S. Dragomir (Ed.), Inequalities for Csiszár f-Divergence in Information Theory, RGMIA Monographs, Victoria University, 2000. (online: http://ajmaa.org/RGMIA/monographs.php/)
[6] T. M. Flett, Differential Analysis, Cambridge University Press, 1980.
[7] E. Hewitt and K. R. Stromberg, Real and Abstract Analysis, Graduate Text in Mathematics 25, Springer-Verlag, Berlin-Heidelberg-New York, 1965.
[8] L. Horváth, Inequalities corresponding to the classical Jensen's inequality, J. Math. Inequal. **3** (2009), 189-200.
[9] L. Horváth, Đ . Pečarić, J. Pečarić, Estimations of f- and Rényi divergences by using a cyclic refinement of the Jensen's inequality, to appear in Bull. Malays. Math. Sci. Soc. 2017. Online publication: doi:10.1007/s40840-017-0526-4
[10] F. Liese and I. Vajda, On divergences and informations in statistics and information theory, IEEE Trans. Inf. Theory **52** (2006), 4394-4412.
[11] B. Mandelbrot, An informational theory of the statistical structure of language, In Jackson, W., editor, Communication Theory, pp. 486-502. Academic Press, New York (1953)
[12] T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[13] I. Sason and S. Verdú, f-Divergence Inequalities, IEEE Trans. Inf. Theory **62** (2016), 5973-6006.
[14] I. Vajda, Theory of Statistical Inference and Information, Boston, MA: Kluwer, 1989.
[15] G. K. Zipf, Selected Studies of the Principle of Relative Frequency in Language. Cambridge, MA: Harvard University Press, 1932.