The Australian Journal of Mathematical Analysis and Applications

AJMAA

Volume 15, Issue 1, Article 1, pp. 1-275, 2018

INEQUALITIES FOR DISCRETE F-DIVERGENCE MEASURES: A SURVEY OF RECENT RESULTS

SILVESTRU SEVER DRAGOMIR[1,2]

Received 19 August, 2017; accepted 15 October, 2017; published 29 January, 2018.

[1]MATHEMATICS, COLLEGE OF ENGINEERING & SCIENCE, VICTORIA UNIVERSITY, PO BOX 14428, MELBOURNE CITY, MC 8001, AUSTRALIA.
sever.dragomir@vu.edu.au
URL: http://rgmia.org/dragomir

[2]DST-NRF CENTRE OF EXCELLENCE, IN THE MATHEMATICAL AND STATISTICAL SCIENCES, SCHOOL OF COMPUTER SCIENCE & APPLIED MATHEMATICS, UNIVERSITY OF THE WITWATERSRAND, PRIVATE BAG 3, JOHANNESBURG 2050, SOUTH AFRICA

ABSTRACT. In this paper we survey some recent results obtained by the author in providing various bounds for the celebrated f-divergence measure for various classes of functions f. Several techniques including inequalities of Jensen and Slater types for convex functions are employed. Bounds in terms of Kullback-Leibler Distance, Hellinger Discrimination and Varation distance are provided. Approximations of the f-divergence measure by the use of the celebrated Ostrowski and Trapezoid inequalities are obtained. More accurate approximation formulae that make use of Taylor's expansion with integral remainder are also surveyed. A comprehensive list of recent papers by several authors related this important concept in information theory is also included as an appendix to the main text.

Key words and phrases: f-divergence, Measure of information, Entropy, Analytic inequalities, Jensen's inequality, Convex functions, Integral inequalities, Ostrowski's inequality, Trapezoid inequality, Taylor's expansion.

2000 *Mathematics Subject Classification.* Primary 94A17, 26D15; Secondary 62B10

ISSN (electronic): 1449-5910

This paper is in final form and no version of it will be submitted for publication elsewhere.

Contents

1. INTRODUCTION

In this introductory chapter we present the definition and some fundamental results for f-divergence measure. Certain examples that are useful in various applications including mathematical statistics, information theory and signal processing are provided as well.

Given a convex function $f : \mathbb{R}_+ \to \mathbb{R}_+$, the *$f$-divergence functional*, or *f-divergence measure*

$$I_f(p,q) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \tag{1.1}$$

was introduced by Csiszár in [**25**], [**23**] as a generalized measure of information, a "distance function" on the set of probability distributions $\mathbb{P}^n$.

The restriction to discrete distributions is only for convenience, similar results hold for more general distributions.

The definition (1.1) can be extended for nonconvex function, however in this case the positivity property of $I_f(p,q)$ is not always assured.

As in Csiszár [**23**], we interpret the following, otherwise undefined expressions, as indicated:

$$f(0) = \lim_{t\to 0+} f(t), \quad 0f\left(\frac{0}{0}\right) = 0,$$
$$0f\left(\frac{a}{0}\right) = \lim_{\varepsilon\to 0+} f\left(\frac{a}{\varepsilon}\right) = a\lim_{t\to\infty}\frac{f(t)}{t}, \quad a > 0.$$

The immediately following results were essentially given by Csiszár and Körner [**26**].

THEOREM 1.1 (Csiszár & Körner, 1981 [**26**]). *If $f : \mathbb{R}_+ \to \mathbb{R}$ is convex, then $I_f(p,q)$ is jointly convex in p and q.*

The following lower bound for the f-divergence functional also holds.

THEOREM 1.2 (Csiszár & Körner, 1981 [**26**]). *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be convex, then for every $p,q \in \mathbb{R}_+^n$, we have the inequality:*

$$I_f(p,q) \geq \sum_{i=1}^{n} q_i f\left(\frac{\sum_{i=1}^{n} p_i}{\sum_{i=1}^{n} q_i}\right). \tag{1.2}$$

If f is strictly convex, equality holds in (1.2) iff

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = \cdots = \frac{p_n}{q_n}. \tag{1.3}$$

COROLLARY 1.3. *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be convex and normalized, i.e.,*

$$f(1) = 0, \tag{1.4}$$

then, for any $p,q \in \mathbb{R}_+^n$ with

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i, \tag{1.5}$$

we have the inequality,

$$I_f(p,q) \geq 0. \tag{1.6}$$

If f is strictly convex, equality holds in (1.6) iff $p_i = q_i$ for all $i \in \{1,\dots,n\}$.

In particular, if p,q are probability vectors, then (1.5) is assured. Corollary 1.3 then shows that, for strictly convex and normalized $f:\mathbb{R}_{+}\rightarrow\mathbb{R}$,

$$I_f(p,q)\geq 0\quad\text{for all } p,q\in\mathbb{P}^n. \tag{1.7}$$

The equality holds in (1.7) iff $p=q$.

These are "distance properties", however, I_f is not a metric since it violates the triangle inequality, and is asymmetric, i.e, for general $p,q\in\mathbb{R}_{+}^{n}$, $I_f(p,q)\neq I_f(q,p)$.

2. Some Examples

In the examples below we obtain, for suitable choices of the kernel f, some of the best known distance functions I_f used in mathematical statistics [**85**] – [**87**], information theory [**12**], [**119**] and signal processing [**79**], [**98**].

Example 2.1. *(Kullback-Leibler) For*

$$f(t):=t\log t,\ t>0 \tag{2.1}$$

the f-divergence is

$$I_f(p,q)=KL(p,q)=\sum_{i=1}^{n}p_i\log\left(\frac{p_i}{q_i}\right), \tag{2.2}$$

called the Kullback-Leibler distance [**94**]-[**97**].

Example 2.2. *(Hellinger) Let*

$$f(t)=\frac{1}{2}\left(1-\sqrt{t}\right)^2,\ t>0. \tag{2.3}$$

Then I_f gives the Hellinger distance [**100**]

$$I_f(p,q)=h^2(p,q)=\frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2, \tag{2.4}$$

which is symmetric.

Example 2.3. *(Renyi) For $\alpha>1$, let*

$$f(t)=t^{\alpha},\ t>0. \tag{2.5}$$

Then

$$I_f(p,q)=D_{\alpha}(p,q)=\sum_{i=1}^{n}p_i^{\alpha}q_i^{1-\alpha}, \tag{2.6}$$

which is the α-order entropy [**117**].

Example 2.4. *(χ^2-distance) Let*

$$f(t)=(t-1)^2,\ t>0. \tag{2.7}$$

Then

$$I_f(p,q) = D_{\chi^2}(p,q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} \tag{2.8}$$
$$= \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 2P_n + Q_n$$
$$\left(= \sum_{i=1}^{n} \frac{p_i^2 - q_i^2}{q_i} \quad \textit{if } P_n = Q_n \right)$$

is the χ^2*-distance between* p *and* q*, where* $P_n = \sum_{i=1}^{n} p_i$ *and* $Q_n = \sum_{i=1}^{n} q_i$.

Finally, we have

EXAMPLE 2.5. *(Variation distance). Let* $f(t) = |t-1|$, $t > 0$*. The corresponding* f*-divergence, called the variation distance, is symmetric,*

$$V(p,q) = \sum_{i=1}^{n} |p_i - q_i|.$$

For other examples of divergence measures, see the paper [**88**] by J. N. Kapur, where further references are given.

For other examples of divergence measures and further references, see [**88**].

CHAPTER 1

Inequalities for f-Divergence

In this chapter we present some Jensen type inequalities with emphasis on some reverses and apply them for the general case of f-divergence measure. We make use of Grüss type inequalities to provide simpler upper bounds and apply the general results for some particular divergence measures of interest.

1. Jensen's Type Inequalities

We start with the following general result:

THEOREM 1.1 (Dragomir, 2003 [**43**]). *Let $f:\mathbb{R}_+\to\mathbb{R}$ be differentiable convex, then for all $p,q\in\mathbb{R}_+^n$,*

$$f'(1)(P_n-Q_n)\le I_f(p,q)-Q_nf(1)\le I_{f'}\left(\frac{p^2}{q},p\right)-I_{f'}(p,q), \tag{1.1}$$

where $P_n:=\sum\limits_{i=1}^n p_i>0$, $Q_n:=\sum\limits_{i=1}^n q_i>0$ and $f':(0,\infty)\to\mathbb{R}$ is the derivative of f.
If f is strictly convex and $p_i,q_i>0$ $(i=1,\dots,n)$, then equality holds in (1.1) iff $p=q$.

PROOF. We follow the proof in [**43**].
As f is differentiable convex on $\mathbb{R}_+$, then we have the inequality,

$$f'(y)(y-x)\ge f(y)-f(x)\ge f'(x)(y-x) \tag{1.2}$$

for all $x,y\in\mathbb{R}_+$.
Choose $y=\frac{p_i}{q_i}$ and $x=1$ in (1.2), to obtain,

$$f'\left(\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-1\right)\ge f\left(\frac{p_i}{q_i}\right)-f(1)\ge f'(1)\left(\frac{p_i}{q_i}-1\right) \tag{1.3}$$

for all $i\in\{1,\dots,n\}$.
Now, if we multiply (1.3) by $q_i\ge 0$ $(i=1,\dots,n)$ and sum over i from 1 to n, we deduce,

$$\sum_{i=1}^n (p_i-q_i)f'\left(\frac{p_i}{q_i}\right)\ge I_f(p,q)-Q_nf(1)\ge f'(1)(P_n-Q_n)$$

and as

$$\sum_{i=1}^n (p_i-q_i)f'\left(\frac{p_i}{q_i}\right)=I_{f'}\left(\frac{p^2}{q},p\right)-I_{f'}(p,q),$$

the inequality in (1.1) is obtained.
The case of equality holds in (1.2) for a strictly convex mapping iff $x=y$ and so the equality holds in (1.1) iff $\frac{p_i}{q_i}=1$ for all $i\in\{1,\dots,n\}$, and the theorem is proved. ∎

REMARK 1.1. In the above theorem, if the differentiability condition is dropped, we can choose instead of $f'(x)$, any number $l=l(x)\in\left[f'_-(x),f'_+(x)\right]$ and the inequality is still valid. This follows by the fact that for the convex mapping $f:\mathbb{R}_+\to\mathbb{R}_+$,

$$l_2(x)(x-y)\ge f(x)-f(y)\ge l_1(y)(x-y),\quad x,y\in(0,\infty);$$

where $l_1(y) \in [f'_-(y), f'_+(y)]$ and $l_2(x) \in [f'_-(x), f'_+(x)]$.

The following corollary is a natural consequence of the above theorem.

COROLLARY 1.2 (Dragomir, 2003 [**43**]). *fLet $f : \mathbb{R}_+ \to \mathbb{R}_+$ be convex and normalized. If $f'(1)(P_n - Q_n) \geq 0$, then we have the positivity inequality,*

$$0 \leq I_f(p,q) \leq I_{f'}\left(\frac{p^2}{q}, p\right) - I_{f'}(p,q). \tag{1.4}$$

The equality holds in (1.4) for a strictly convex mapping f iff $p = q$.

THEOREM 1.3 (Dragomir, 2003 [**43**]). *Assume that f is differentiable convex on $(0,\infty)$. If $p^{(j)}, q^{(j)}$ $(j = 1,2)$ are probability distributions, then for all $\lambda \in [0,1]$,*

$$\begin{aligned} 0 &\leq \lambda I_f\left(p^{(1)}, q^{(1)}\right) + (1-\lambda) I_f\left(p^{(2)}, q^{(2)}\right) \\ &\quad - I_f\left(\lambda p^{(1)} + (1-\lambda) q^{(1)}, \lambda p^{(2)} + (1-\lambda) q^{(2)}\right) \\ &\leq \lambda(1-\lambda) \sum_{i=1}^{n} \frac{\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} \left[f'\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) - f'\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right], \end{aligned} \tag{1.5}$$

where f' is the derivative of f.

PROOF. We follow the proof of [**43**].
Using (1.2),

$$\begin{aligned} &f'\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right) \\ &\geq f\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) \\ &\geq f'\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right) \end{aligned} \tag{1.6}$$

and

$$\begin{aligned} &f'\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right) \\ &\geq f\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right) \\ &\geq f'\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right). \end{aligned} \tag{1.7}$$

Multiplying (1.6) by $\lambda q_i^{(1)}$, (1.7) by $(1-\lambda)\, q_i^{(2)}$ and adding gives:-

$$
\begin{aligned}
(1.8)\qquad & \sum_{i=1}^{n} f'\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}}\right)\left[\lambda q_i^{(1)}\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right)\right. \\
&\qquad \left. + (1-\lambda)\, q_i^{(2)}\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right)\right] \\
& \geq I_f\left(\lambda p^{(1)} + (1-\lambda)\, p^{(2)}, \lambda q^{(1)} + (1-\lambda)\, q^{(2)}\right) \\
&\qquad - \lambda I_f\left(p^{(1)}, q^{(1)}\right) - (1-\lambda)\, I_f\left(p^{(2)}, q^{(2)}\right) \\
& \geq \sum_{i=1}^{n}\left[\lambda q_i^{(1)} f'\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right)\right. \\
&\qquad \left. + (1-\lambda)\, q_i^{(2)} f'\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right)\right].
\end{aligned}
$$

However,

$$
\begin{aligned}
& \lambda q_i^{(1)}\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right) \\
&\qquad\qquad + (1-\lambda)\, q_i^{(2)}\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right) \\
&\qquad\qquad = -\frac{\lambda(1-\lambda)\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} + \frac{\lambda(1-\lambda)\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} = 0,
\end{aligned}
$$

which shows that the first term in (1.8) is zero.

In addition,

$$
\lambda q_i^{(1)}\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right) = -\frac{\lambda(1-\lambda)\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}},
$$

and

$$
(1-\lambda)\, q_i^{(2)}\left(\frac{\lambda p_i^{(1)} + (1-\lambda)\, p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}} - \frac{p_i^{(2)}}{q_i^{(2)}}\right) = -\frac{\lambda(1-\lambda)\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}},
$$

and so the second term in (1.4) is,

$$
-\lambda(1-\lambda)\sum_{i=1}^{n}\frac{\begin{vmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{vmatrix}}{\lambda q_i^{(1)} + (1-\lambda)\, q_i^{(2)}}\left[f'\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) - f'\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right],
$$

which proves the theorem. ∎

REMARK 1.2. The first inequality in (1.5) is actually the joint convexity property of $I_f(\cdot,\cdot)$ which has been proven here in a different manner from that in [**26**].

We have the following reverse of Jensen's discrete inequality established in 1994 in [**62**] by Dragomir and Ionescu.

LEMMA 1.4 (Dragomir & Ionescu, 1994 [**62**]). *Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable convex mapping on I. If $x_i \in \mathring{I}$, the interior of I, $t_i \geq 0$ with $T_n := \sum_{i=1}^{n} t_i > 0 \quad (i = 1, \ldots, n)$, then we have the inequality,*

$$0 \leq \frac{1}{T_n} \sum_{i=1}^{n} t_i f(x_i) - f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i\right) \leq \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i f'(x_i) - \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \cdot \frac{1}{T_n} \sum_{i=1}^{n} t_i f'(x_i); \tag{1.9}$$

where $f' : \mathring{I} \to \mathbb{R}$ is the derivative of f on $\mathring{I}$.
If $t_i > 0$ $(i = 1, \ldots, n)$ and the mapping f is strictly convex on $\mathring{I}$, then the case of equality holds in (1.9) iff $x_1 = x_2 = \cdots = x_n$.

PROOF. For the sake of completeness, we provide here a short proof.
As $f : I \to \mathbb{R}$ is convex on $\mathring{I}$, then we have,

$$f(x) - f(y) \geq f'(y)(x - y) \quad \text{for all } x, y \in \mathring{I}. \tag{1.10}$$

In (1.10), choose $x = \frac{1}{T_n} \sum\limits_{i=1}^{n} t_i x_i \in \mathring{I}$, $y = x_j,\ j = 1, \ldots, n$ to obtain,

$$f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i\right) - f(x_j) \geq f'(x_j)\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i - x_j\right) \tag{1.11}$$

for all $j \in \{1, \ldots, n\}$.
Multiplying (1.11) by $t_j \geq 0$ and summing over j from 0 to n, gives

$$\sum_{j=1}^{n} t_j \left[f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i\right) - f(x_j)\right] \geq \sum_{j=1}^{n} t_j f'(x_j)\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i - x_j\right). \tag{1.12}$$

A simple calculation shows that,

$$\sum_{j=1}^{n} t_j \left[f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i\right) - f(x_j)\right] = T_n f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i\right) - \sum_{i=1}^{n} t_i f(x_i)$$

and

$$\sum_{j=1}^{n} t_j f'(x_j)\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i - x_j\right) = \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \sum_{i=1}^{n} t_i f'(x_i) - \sum_{i=1}^{n} t_i x_i f'(x_i)$$

and then, by (1.12),

$$T_n f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i\right) - \sum_{i=1}^{n} t_i f(x_i) \geq \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \sum_{i=1}^{n} t_i f'(x_i) - \sum_{i=1}^{n} t_i x_i f'(x_i).$$

Dividing by $T_n > 0$, we obtain the second inequality in (1.9).
If f is strictly convex, then the equality holds in (1.10) iff $x = y$. Using this and an obvious argument, the equality holds in (1.9) iff $x_1 = \cdots = x_n$. ∎

REMARK 1.3. If, in the above lemma we drop the differentiability condition, and choose, instead of $f'(x_i)$, any number $l = l(x_i) \in [f'_-(x_i), f'_+(x_i)]$, the inequality still remains valid. This follows by virtue of the fact that for a convex mapping $f : \mathbb{R}_+ \to \mathbb{R}_+$, we have,

$$f(x) - f(y) \geq l(y)(x - y) \qquad \text{for } x, y \in \mathbb{R}_+,$$

where $l(y) \in [f'_-(y), f'_+(y)]$.

REMARK 1.4. For an extension of this theorem to convex mappings of several variables see [**70**] by Dragomir and Goh where further applications in Information Theory for Shannon's Entropy, Mutual Information, Conditional Entropy, etc. are given. An integral version of this result can be found in [**74**] by Dragomir and Goh where further applications for the continuous case of Shannon's Entropy have been given. Extensions of the above results for convex mappings defined on convex sets in linear spaces, and particularly in normed spaces, can be found in the Ph.D. dissertation [**103**] by M. Matić where other applications in Information Theory for Shannon's Entropy have been considered.

The following reverse inequality for the f-divergence also holds [**43**].

THEOREM 1.5 (Dragomir, 2003 [**43**]). *Let f, p and q be as in Theorem 1.1, then we have the inequality:*

$$0 \leq I_f(p,q) - Q_n f\left(\frac{P_n}{Q_n}\right) \leq I_{f'}\left(\frac{p^2}{q}, p\right) - \frac{P_n}{Q_n} I_{f'}(p,q), \tag{1.13}$$

where $Q_n := \sum_{i=1}^n q_i > 0$, $P_n := \sum_{i=1}^n p_i > 0$. If f is strictly convex, and p_i, $q_i > 0$, $(i = 1, \ldots, n)$, then the equality holds in (1.13) iff $\frac{p_1}{q_1} = \cdots = \frac{p_n}{q_n}$.

PROOF. In Lemma 1.4 choose $f = f$, $t_i := q_i$ and $x_i = \frac{p_i}{q_i}$ to obtain,

$$\begin{aligned} 0 &\leq \frac{1}{Q_n}\sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right) - f\left(\frac{P_n}{Q_n}\right) \\ &\leq \frac{1}{Q_n}\sum_{i=1}^n p_i I_{f'}\left(\frac{p_i}{q_i}\right) - \frac{1}{Q_n}\sum_{i=1}^n p_i \cdot \frac{1}{Q_n}\sum_{i=1}^n q_i I_{f'}\left(\frac{p_i}{q_i}\right), \end{aligned}$$

which is equivalent to (1.13). The case of equality is obvious. ∎

The corollary below follows as a natural consequence.

COROLLARY 1.6. *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be differentiable convex and normalized. For any $p, q \in \mathbb{R}^n_+$ with $P_n = Q_n$, we have then the reverse of the positivity inequality (1.7),*

$$0 \leq I_f(p,q) \leq I_{f'}\left(\frac{p^2}{q}, p\right) - I_{f'}(p,q). \tag{1.14}$$

The equality holds in (1.14) for a strictly convex mapping f iff $p = q$.

2. APPLICATIONS FOR SOME PARTICULAR f-DIVERGENCES

Consider the convex function $f(t) = -\log t$, $t > 0$. For this function we have the f-divergence,

$$I_f(p,q) = \sum_{i=1}^n q_i \left[-\log\left(\frac{p_i}{q_i}\right)\right] = \sum_{i=1}^n q_i \log\left(\frac{q_i}{p_i}\right) = KL(q,p). \tag{2.1}$$

PROPOSITION 2.1. *Let* $p,q\in\mathbb{R}^n$*, then,*

$$Q_n - P_n \le KL(q,p) \le \sum_{i=1}^{n} \frac{q_i^2}{p_i} - Q_n. \tag{2.2}$$

The case of equality holds iff $p=q$.

PROOF. Since $f(t)=-\log t$, then $f'(t)=-\frac{1}{t}$, $t>0$. We have,

$$I_{f'}\left(\frac{p^2}{q},p\right)=\sum_{i=1}^{n}p_i\cdot\left[-\frac{1}{\left(\frac{p_i^2}{q_i}\right)\cdot\frac{1}{p_i}}\right]=-Q_n,$$
$$I_{f'}(p,q)=\sum_{i=1}^{n}q_i\cdot\left[-\frac{1}{\frac{p_i}{q_i}}\right]=-\sum_{i=1}^{n}\frac{q_i^2}{p_i},$$

and then, from (1.1), we obtain,

$$-(P_n-Q_n)\le KL(q,p)\le -Q_n+\sum_{i=1}^{n}\frac{q_i^2}{p_i},$$

which is the desired inequality (2.2).

The case of equality is obvious, taking into account that $-\log$ is a strictly convex mapping on $(0,\infty)$. ∎

The following result for the Kullback-Leibler distance also holds.

PROPOSITION 2.2. *Let* $p,q\in\mathbb{R}^n$*, then,*

$$P_n-Q_n\le KL(p,q)\le P_n-Q_n+KL(q,p)-KL\left(p,\frac{p^2}{q}\right). \tag{2.3}$$

The case of equality holds iff $p=q$.

PROOF. As $f(t)=t\log(t)$, then $f'(t)=\log t+1$. We have, then,

$$I_f(p,q)=KL(p,q),$$
$$I_{f'}\left(\frac{p^2}{q},p\right)=I_{\log(\cdot)+1}\left(\frac{p^2}{q},p\right)=P_n+I_{\log(\cdot)}\left(\frac{p^2}{q},p\right).$$

Since $I_{-\log(\cdot)}(p,q)=KL(q,p)$ (see (2.1)), then,

$$I_{\log(\cdot)}\left(\frac{p^2}{q},p\right)=-KL\left(p,\frac{p^2}{q}\right).$$

In addition,

$$\begin{aligned}I_{f'}(p,q)&=I_{\log(\cdot)+1}(p,q)=Q_n+I_{\log(\cdot)}(p,q)\\&=Q_n-KL(q,p)\end{aligned}$$

and so, by (1.1), we can state that,

$$P_n-Q_n\le KL(p,q)\le P_n-Q_n-KL\left(p,\frac{p^2}{q}\right)Q_n+KL(q,p)$$

and the inequality (2.3) is obtained.

The case of equality holds from the fact that the mapping $f(t)=t\log t$ is strictly convex on $(0,\infty)$. ∎

It is known that Rényi's entropy is actually the f-divergence for the convex mapping $f(t) = t^{\alpha}$, $\alpha > 1$, $t > 0$).

PROPOSITION 2.3. *Let $p, q \in \mathbb{R}^n_+$, then,*

$$\alpha (P_n - Q_n) \le D_\alpha (p,q) - Q_n \le \alpha \left[D_\alpha (p,q) - D_\alpha \left(q^{\frac{2-\alpha}{\alpha}}, p^{-1} \right) \right]. \tag{2.4}$$

The case of equality holds iff $p = q$.

PROOF. Since $f(t) = t^{\alpha}$, then $f'(t) = \alpha t^{\alpha - 1}$.

We have

$$\begin{aligned} I_{f'}\left(\frac{p^2}{q}, p\right) &= \sum_{i=1}^n p_i \left[\alpha \cdot \left(\frac{p_i^2}{q_i p_i} \right)^{\alpha - 1} \right] \\ &= \alpha \sum_{i=1}^n p_i \left(\frac{p_i}{q_i} \right)^{\alpha-1} = \alpha \sum_{i=1}^n q_i^{1-\alpha} p_i^{\alpha} = \alpha D_\alpha (p,q) \end{aligned}$$

and

$$\begin{aligned} I_{f'}(p,q) &= \sum_{i=1}^n q_i \left[\alpha \cdot \left(\frac{p_i}{q_i} \right)^{\alpha-1} \right] \\ &= \alpha \sum_{i=1}^n p_i^{\alpha-1} q_i^{2-\alpha} = \alpha D_\alpha \left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p} \right). \end{aligned}$$

Using (1.1),

$$\alpha (P_n - Q_n) \le D_\alpha (p,q) - Q_n \le \alpha \left[D_\alpha (p,q) - D_\alpha \left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p} \right) \right]$$

and (2.4) is proved.

The case of equality holds since the mapping $f(t) = t^{\alpha}$ is strictly convex on $(0, \infty)$ for $\alpha > 1$. ∎

PROPOSITION 2.4. *Let $p, q \in \mathbb{R}^n_+$, then,*

$$0 \le h^2 (p,q) \le \frac{1}{2} [P_n - Q_n] + \frac{1}{2} \left[\sum_{i=1}^n q_i \left(\sqrt{\frac{q_i}{p_i}} - \sqrt{\frac{p_i}{q_i}} \right) \right]. \tag{2.5}$$

The equality holds iff $p = q$.

PROOF. As $f(t) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$, we have $f'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}}$ and $f''(t) = \frac{1}{4} \cdot \frac{1}{\sqrt{t^3}} > 0$ $(t \in (0,\infty))$ which shows that f is indeed strictly convex on $(0, \infty)$.

We also have,

$$\begin{aligned} I_f(p,q) &= h^2(p,q), \\ I_{f'}\left(\frac{p^2}{q}, p\right) &= \sum_{i=1}^n p_i \left[\frac{1}{2} - \frac{1}{2\sqrt{\frac{p_i^2}{q_i p_i}}} \right] \\ &= \frac{1}{2} P_n - \frac{1}{2} \sum_{i=1}^n \sqrt{p_i q_i} = \frac{1}{2} \left[P_n - \sum_{i=1}^n \sqrt{p_i q_i} \right] \\ I_{f'}(p,q) &= \sum_{i=1}^n q_i \left[\frac{1}{2} - \frac{1}{2\sqrt{\frac{p_i}{q_i}}} \right] = \frac{1}{2} \left[Q_n - \sum_{i=1}^n q_i \sqrt{\frac{q_i}{p_i}} \right] \end{aligned}$$

and as $f'(1) = 0$ and $f(1) = 0$, then, by (1.1) applied for f as above, we deduce (2.5). The case of equality is obvious by the strict convexity of f. ∎

Consider *Bhattacharyya's distance* (see for example [**88**]),

$$B(p,q) = \sum_{i=1}^{n} \sqrt{p_i q_i},$$

where $p, q \in \mathbb{R}_+^n$.

We know that for the convex mapping $f(t) = -\sqrt{t}$,

$$I_f(p,q) = -\sum_{i=1}^{n} q_i \sqrt{\frac{p_i}{q_i}} = -B(p,q).$$

PROPOSITION 2.5. *Let $p, q \in \mathbb{R}_+^n$, then,*

$$\frac{1}{2}(Q_n - P_n) \le Q_n - B(p,q) \le \frac{1}{2}\sum_{i=1}^{n} q_i \left(\sqrt{\frac{q_i}{p_i}} - \sqrt{\frac{p_i}{q_i}}\right). \tag{2.6}$$

The case of equality holds iff $p = q$.

PROOF. As $f(1) = -\sqrt{t}$, $t > 0$, then $f'(t) = -\frac{1}{2\sqrt{t}}$ and $f''(t) = \frac{1}{4\sqrt{t^3}}$, $t > 0$, which also shows that $f(\cdot)$ is strictly convex on $(0, \infty)$. We also have,

$$I_{f'}\left(\frac{p^2}{q}, p\right) = \sum_{i=1}^{n} p_i \left[-\frac{1}{2\sqrt{\frac{p_i^2}{q_i p_i}}}\right] = -\frac{1}{2}\sum_{i=1}^{n} \sqrt{p_i q_i} = -\frac{1}{2}B(p,q),$$

$$I_{f'}(p,q) = -\frac{1}{2}\sum_{i=1}^{n} q_i \frac{1}{\sqrt{\frac{p_i}{q_i}}} = -\frac{1}{2}\sum_{i=1}^{n} q_i \sqrt{\frac{q_i}{p_i}}$$

and as $f'(1) = -\frac{1}{2}$, $f(1) = -1$, then by (1.1) applied for the mapping f as defined above, we deduce (2.6).

The case of equality is obvious by the strict convexity of f. ∎

We continue now with some particular inequalities which may be obtained from Theorem 1.5.

PROPOSITION 2.6. *Let $p, q \in \mathbb{R}_+^n$, then,*

$$0 \le KL(q,p) - Q_n \log\left(\frac{Q_n}{P_n}\right) \le \frac{1}{Q_n}\left[P_n \sum_{i=1}^{n} \frac{q_i^2}{p_i} - Q_n^2\right] \tag{2.7}$$

$$= \frac{1}{2Q_n}\sum_{i,j=1}^{n} p_i p_j \left(\frac{q_i}{p_i} - \frac{q_j}{p_j}\right)^2,$$

with equality iff $\frac{p_1}{q_1} = \cdots = \frac{p_n}{q_n}$.

PROOF. If $f(t) = -\log t$, then $f'(t) = -\frac{1}{t}$, $t > 0$. We have,

$$I_{f'}\left(\frac{p^2}{q}, p\right) = \sum_{i=1}^{n} p_i \cdot \left[\frac{1}{\left(\frac{p_i^2}{q_i}\right) \cdot \frac{1}{p_i}}\right] = -Q_n,$$

$$I_{f'}(p,q) = \sum_{i=1}^{n} q_i \cdot \left[-\frac{1}{\frac{p_i}{q_i}}\right] = -\sum_{i=1}^{n} \frac{q_i^2}{p_i}$$

and from (1.13),

$$0 \le KL(q,p) + Q_n \log\left(\frac{P_n}{Q_n}\right) \le -Q_n + \frac{P_n}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i} = \frac{1}{Q_n}\left[P_n\sum_{i=1}^{n}\frac{q_i^2}{p_i} - Q_n^2\right],$$

i.e.,

$$0 \le KL(q,p) - Q_n \log\left(\frac{Q_n}{P_n}\right) \le \frac{1}{Q_n}\left[P_n\sum_{i=1}^{n}\frac{q_i^2}{p_i} - Q_n^2\right].$$

On the other hand, we observe that,

$$\begin{aligned}\frac{1}{2}\sum_{i,j=1}^{n}p_ip_j\left(\frac{q_i}{p_i}-\frac{q_j}{p_j}\right)^2 &= \frac{1}{2}\sum_{i,j=1}^{n}p_ip_j\left(\frac{q_i^2}{p_i^2}-2\frac{q_iq_j}{p_ip_j}+\frac{q_j^2}{p_j^2}\right)\\ &= \frac{1}{2}\left[\sum_{i,j=1}^{n}p_j\frac{q_i^2}{p_i} - 2\sum_{i,j=1}^{n}q_iq_j + \sum_{i,j=1}^{n}p_i\frac{q_j^2}{p_j}\right]\\ &= P_n\sum_{i=1}^{n}\frac{q_i^2}{p_i} - Q_n^2\end{aligned}$$

and the inequality (2.7) is proved.

The case of equality follows by the strict convexity of the mapping $-\log$. ∎

COROLLARY 2.7. *If $P_n = Q_n$ in (2.7),*

$$0 \le KL(q,p) \le \sum_{i=1}^{n}\frac{q_i^2-p_i^2}{p_i} = \frac{1}{2Q_n}\sum_{i,j=1}^{n}p_ip_j\left(\frac{q_i}{p_i}-\frac{q_j}{p_j}\right)^2, \tag{2.8}$$

with equality iff $p = q$.

REMARK 2.1. We know that the χ^2-distance between p and q is

$$D_{\chi^2}(p,q) := \sum_{i=1}^{n}\frac{(p_i-q_i)^2}{q_i} = \left(\sum_{i=1}^{n}\frac{p_i^2-q_i^2}{q_i} \text{ if } P_n = Q_n\right).$$

As

$$D_{\chi^2}(q,p) = \sum_{i=1}^{n}\frac{(q_i-p_i)^2}{p_i} = \left(\sum_{i=1}^{n}\frac{q_i^2-p_i^2}{p_i} \text{ if } P_n = Q_n\right),$$

then (2.8) can be rewritten as,

$$0 \le KL(q,p) \le D_{\chi^2}(q,p). \tag{2.9}$$

COROLLARY 2.8. *Let p and q be two probability distributions, then,*

$$0 \le KL(q,p) \le D_{\chi^2}(q,p), \tag{2.10}$$

with equality iff $p = q$.

REMARK 2.2. For a direct proof of (2.10) see [**76**] where further bounds are also given.

PROPOSITION 2.9. *Let $p, q \in \mathbb{R}_+^n$, then,*

$$0 \le D_\alpha(p,q) - P_n^\alpha Q_n^{1-\alpha} \le \frac{\alpha}{Q_n}\left[Q_nD_\alpha(p,q) - P_nD_\alpha\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right)\right], \tag{2.11}$$

with equality iff $\frac{p_1}{q_1} = \cdots = \frac{p_n}{q_n}$.

PROOF. If $f(t) = t^{\alpha}$, then $f'(t) = \alpha t^{\alpha-1}$.
We have,

$$I_{f'}\left(\frac{p^2}{q}, p\right) = \sum_{i=1}^{n} p_i \cdot \left[\alpha\left(\frac{p_i^2}{q_i p_i}\right)^{\alpha-1}\right] = \alpha\sum_{i=1}^{n} p_i\left(\frac{p_i}{q_i}\right)^{\alpha-1} = \alpha\sum_{i=1}^{n} q_i^{1-\alpha}p_i^{\alpha} = \alpha D_{\alpha}(p,q) \tag{2.12}$$

and

$$I_{f'}(p,q) = \sum_{i=1}^{n} q_i\left[\alpha\left(\frac{p_i}{q_i}\right)^{\alpha-1}\right] = \sum_{i=1}^{n} p_i^{\alpha-1}q_i^{2-\alpha} = \alpha D_{\alpha}\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right). \tag{2.13}$$

Using (1.13),

$$0 \le D_{\alpha}(p,q) - Q_n\left(\frac{P_n}{Q_n}\right)^{\alpha} \le \alpha D_{\alpha}(p,q) - \alpha\frac{P_n}{Q_n}D_{\alpha}\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right) = \frac{\alpha}{Q_n}\left[Q_n D_{\alpha}(p,q) - P_n D_{\alpha}\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right)\right]$$

and the inequality (2.11) is obtained.

The case of equality follows by the fact that the mapping $f(t) = t^{\alpha}$ ($\alpha > 1,\ t > 0$) is strictly convex on $(0,\infty)$. ∎

COROLLARY 2.10. *If $P_n = Q_n$ in (2.11),*

$$0 \le D_{\alpha}(p,q) - P_n \le \alpha\left[D_{\alpha}(p,q) - D_{\alpha}\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right)\right], \tag{2.14}$$

with equality iff $p = q$.
In particular, if p, q are probability distributions, then,

$$0 \le D_{\alpha}(p,q) - 1 \le \alpha\left[D_{\alpha}(p,q) - D_{\alpha}\left(q^{\frac{2-\alpha}{\alpha}}, \frac{1}{p}\right)\right], \tag{2.15}$$

with equality iff $p = q$.

PROPOSITION 2.11. *Let $p, q \in \mathbb{R}_+^n$, then,*

$$0 \le \sqrt{Q_n P_n} - B(p,q) \le \frac{1}{2}\left[\frac{P_n}{Q_n}\sum_{i=1}^{n} q_i\sqrt{\frac{q_i}{p_i}} - B(p,q)\right], \tag{2.16}$$

with equality iff $\frac{p_1}{q_1} = \cdots = \frac{p_n}{q_n}$.

PROOF. If $f(t) = -\sqrt{t}$, then $f'(t) = -\frac{1}{2}\cdot\frac{1}{\sqrt{t}}$, $t > 0$. It follows, then, that,

$$I_{f'}\left(\frac{p^2}{q}, p\right) = \sum_{i=1}^{n} p_i\left(-\frac{1}{2}\cdot\frac{1}{\sqrt{\frac{p_i^2}{q_i p_i}}}\right) = -\frac{1}{2}\sum_{i=1}^{n}\sqrt{q_i p_i} = -\frac{1}{2}B(p,q),$$

$$I_{f'}(p,q) = \sum_{i=1}^{n} q_i\left(-\frac{1}{2}\cdot\frac{1}{\sqrt{\frac{p_i}{q_i}}}\right) = -\frac{1}{2}\sum_{i=1}^{n} q_i\sqrt{\frac{q_i}{p_i}}$$

and, from (1.13),

$$0 \le -B(p,q) + Q_n\sqrt{\frac{P_n}{Q_n}} \le \frac{1}{2}B(p,q) + \frac{1}{2}\frac{P_n}{Q_n}\sum_{i=1}^{n} q_i\sqrt{\frac{q_i}{p_i}},$$

which is equivalent to (2.16).

The case of equality follows by the strict convexity of the mapping f. ∎

REMARK 2.3. The second inequality in (2.16) is equivalent to

$$\sqrt{P_n Q_n} \le \frac{1}{2}\left[\frac{P_n}{Q_n}\sum_{i=1}^{n} q_i\sqrt{\frac{q_i}{p_i}} + B(p,q)\right], \tag{2.17}$$

with equality iff $\frac{p_1}{q_1} = \cdots = \frac{p_n}{q_n}$.

COROLLARY 2.12. *If* $P_n = Q_n$ *in (2.16), then,*

$$0 \le P_n - B(p,q) \le \frac{1}{2}\left[\sum_{i=1}^{n} q_i\sqrt{\frac{q_i}{p_i}} - B(p,q)\right],$$

with equality iff $p = q$.

Another important divergence measure in Information Theory is the *J-divergence* defined by,

$$J(p,q) = \sum_{i=1}^{n}(p_i - q_i)\log\left(\frac{p_i}{q_i}\right).$$

Note that the mapping $f(t) := (t-1)\ln t$, $t \in (0,\infty)$, has the derivatives

$$f'(t) = \ln t - \frac{1}{t} + 1, \quad t > 0,\ f''(t) = \frac{t+1}{t^2} > 0 \quad \text{for } t > 0,$$

which shows that f is convex on $(0,1)$ and

$$I_f(p,q) = \sum_{i=1}^{n} q_i\left(\frac{p_i}{q_i} - 1\right)\ln\left(\frac{p_i}{q_i}\right) = J(p,q).$$

PROPOSITION 2.13. *Let* $p, q \in \mathbb{R}_+^n$, *then,*

$$\begin{aligned} 0 &\le J(p,q) - (P_n - Q_n)\ln\left(\frac{P_n}{Q_n}\right) \\ &\le KL(p,q) + \frac{P_n}{Q_n}KL(q,p) + \frac{P_n}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i} - Q_n \end{aligned} \tag{2.18}$$

with equality iff $\frac{p_1}{q_1} = \cdots = \frac{p_n}{q_n}$.

PROOF. We have

$$\begin{aligned} I_{f'}\left(\frac{p^2}{q}, p\right) &= \sum_{i=1}^{n} p_i\left[\ln\left(\frac{p_i^2}{q_i p_i}\right) - \frac{q_i p_i}{p_i^2} + 1\right] \\ &= \sum_{i=1}^{n} p_i\ln\left(\frac{p_i}{q_i}\right) - Q_n + P_n = KL(p,q) - Q_n + P_n, \end{aligned}$$

$$I_{f'}(p,q) = \sum_{i=1}^{n} q_i\left[\ln\left(\frac{p_i}{q_i}\right) - \frac{q_i}{p_i} + 1\right] = -KL(q,p) - \sum_{i=1}^{n}\frac{q_i^2}{p_i} + Q_n$$

and then, from (1.13), we can state that,

$$\begin{aligned}0 &\le J(p,q) - Q_n\left(\frac{P_n}{Q_n}-1\right)\ln\left(\frac{P_n}{Q_n}\right)\\ &\le KL(p,q) - Q_n + P_n - \frac{P_n}{Q_n}\left[-KL(q,p) - \sum_{i=1}^{n}\frac{q_i^2}{p_i} + Q_n\right]\\ &= KL(p,q) + P_n + \frac{P_n}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i} - P_n - Q_n\\ &= KL(p,q) + \frac{P_n}{Q_n}KL(q,p) + \frac{P_n}{Q_n}\sum_{i=1}^{n}\frac{q_i^2}{p_i} - Q_n\end{aligned}$$

and the inequality (2.18) is obtained. ∎

3. Further Bounds for the Case when $P_n = Q_n$

The following inequality of the Grüss type is known in the literature as the Biernacki, Pidek and Ryll-Nardzewski inequality (see for example [**110**]).

LEMMA 3.1 (Biernacki, Pidek & Ryll-Nardzewski). *Let a_i, b_i $(i = 1,\ldots,n)$ be real numbers such that*

$$a \le a_i \le A,\; b \le b_i \le B \;\text{ for all } i \in \{1,\ldots,n\}, \tag{3.1}$$

then,

$$\left|\frac{1}{n}\sum_{i=1}^{n}a_ib_i - \frac{1}{n^2}\sum_{i=1}^{n}a_i\sum_{i=1}^{n}b_i\right| \le \frac{1}{n}\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(A-a)(B-b), \tag{3.2}$$

where $[x]$ denotes the integer part of x.

The following inequality holds [**43**] as a consequence.

THEOREM 3.2 (Dragomir, 2003 [**43**]). *Let $f:\mathbb{R}_+\to\mathbb{R}_+$ be differentiable convex. If $p,q\in\mathbb{R}_+^n$ are such that $P_n = Q_n$ and*

$$m \le p_i - q_i \le M, \quad i = 1,\ldots,n \tag{3.3}$$

$$0 < r \le \frac{p_i}{q_i} \le R < \infty, \quad i = 1,\ldots,n, \tag{3.4}$$

then we have the inequality,

$$0 \le I_f(p,q) - Q_n f(1) \le \left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\left(f'(R)-f'(r)\right). \tag{3.5}$$

PROOF. From (1.1),

$$0 \le I_f(p,q) - Q_n f(1) \le \sum_{i=1}^{n}(p_i-q_i)\,f'\left(\frac{p_i}{q_i}\right). \tag{3.6}$$

Applying (3.2),

$$\left|\frac{1}{n}\sum_{i=1}^{n}(p_i-q_i)f'\left(\frac{p_i}{q_i}\right)-\frac{1}{n^2}\sum_{i=1}^{n}(p_i-q_i)\sum_{i=1}^{n}f'\left(\frac{p_i}{q_i}\right)\right| \le \frac{1}{n}\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\left(f'(R)-f'(r)\right), \tag{3.7}$$

as the mapping f' is monotonic nondecreasing, and then,

$$f'(r)\le f'\left(\frac{p_i}{q_i}\right)\le f'(R) \quad \text{for all } i\in\{1,\dots,n\}.$$

As $\sum_{i=1}^{n}(p_i-q_i)=0$, we deduce, using (3.6) and (3.7), the desired result (3.5). ∎

The following inequalities for particular distances are valid.

(1) If $p,q\in\mathbb{R}_n^+$ are such that the conditions (3.3) and (3.4) hold, then,

$$0\le KL(q,p)\le\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\frac{R-r}{rR}, \tag{3.8}$$

and

$$0\le KL(p,q)\le\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\left[\log\left(\frac{R}{r}\right)\right]. \tag{3.9}$$

(2) If p,q are as in (3.3) and (3.4), $(\alpha\ge 1)$ then,

$$0\le D_\alpha(p,q)-Q_n\le\alpha\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\left(R^{\alpha-1}-r^{\alpha-1}\right). \tag{3.10}$$

(3) If p,q are as in (3.3) and (3.4),

$$0\le h^2(p,q)\le\frac{1}{2}\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\cdot\left[\frac{n}{2}\right]\right)(M-m)\frac{\sqrt{R}-\sqrt{r}}{\sqrt{rR}}. \tag{3.11}$$

(4) Under the above assumptions for p and q,

$$0\le Q_n-B(p,q)\le\frac{1}{2}\left[\frac{n}{2}\right]\left(1-\frac{1}{n}\left[\frac{n}{2}\right]\right)(M-m)\frac{\sqrt{R}-\sqrt{r}}{\sqrt{rR}}. \tag{3.12}$$

The following is a Grüss weighted inequality.

LEMMA 3.3. *Assume that* a_i, b_i $(i=1,\dots,n)$ *are as in Lemma 3.1. If* $q_i\ge 0$, $\sum_{i=1}^{n}q_i=1$, *then we have the inequality,*

$$\left|\sum_{i=1}^{n}q_ia_ib_i-\sum_{i=1}^{n}q_ia_i\sum_{i=1}^{n}q_ib_i\right|\le\frac{1}{4}(A-a)(B-b). \tag{3.13}$$

Using this we may prove the following reverse inequality as well [**43**].

THEOREM 3.4 (Dragomir, 2003 [**43**]). *Let* $f:\mathbb{R}_+\to\mathbb{R}_+$ *be differentiable convex. If* $p,q\in\mathbb{R}_+^n$ *are such that* $P_n=Q_n$ *and*

$$0<r\le\frac{p_i}{q_i}\le R<\infty,\quad i=1,\dots,n, \tag{3.14}$$

then we have the inequality,

$$0\le I_f(p,q)-Q_nf(1)\le\frac{1}{4}(R-r)\left[f'(R)-f'(r)\right]. \tag{3.15}$$

PROOF. From (1.1),

$$0 \le I_f(p,q) - Q_n f(1) \le \sum_{i=1}^{n} (p_i - q_i) f'\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right) f'\left(\frac{p_i}{q_i}\right). \tag{3.16}$$

As $f'(\cdot)$ is monotonic nondecreasing, then,

$$f'(r) \le f'\left(\frac{p_i}{q_i}\right) \le f'(R) \quad \text{for all } i \in \{1,\dots,n\}.$$

Applying (3.13) for $a_i = \frac{p_i}{q_i} - 1$, $b_i = f'\left(\frac{p_i}{q_i}\right)$, we obtain

$$\left|\sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right) f'\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right) \sum_{i=1}^{n} q_i f'\left(\frac{p_i}{q_i}\right)\right| \le \frac{1}{4}(R-r)\left[f'(R) - f'(r)\right] \tag{3.17}$$

and as

$$\sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right) = 0,$$

then, by (3.16) and (3.17) we deduce (3.15). ∎

The following inequalities for particular distances are valid.

(1) If p, q are such that $P_n = Q_n$ and (3.14) holds, then,

$$0 \le KL(q,p) \le \frac{(R-r)^2}{4rR} \tag{3.18}$$

and

$$0 \le KL(q,p) \le \frac{1}{4}(R-r)^2 \ln\left(\frac{R}{r}\right). \tag{3.19}$$

(2) With the same assumptions for p, q, we have,

$$0 \le D_\alpha(p,q) - Q_n \le \frac{\alpha}{4}(R-r)\left(R^{\alpha-1} - r^{\alpha-1}\right) \quad (\alpha \ge 1); \tag{3.20}$$

$$0 \le h^2(p,q) \le \frac{1}{8}(R-r)\frac{\sqrt{R} - \sqrt{r}}{\sqrt{Rr}} \tag{3.21}$$

and

$$0 \le Q_n - B(p,q) \le \frac{1}{8}(R-r)\frac{\sqrt{R} - \sqrt{r}}{\sqrt{Rr}}. \tag{3.22}$$

REMARK 3.1. Any other Grüss type inequality can be used to provide different bounds for the difference

$$\Delta := \sum_{i=1}^{n} (p_i - q_i) f'\left(\frac{p_i}{q_i}\right).$$

We omit the details.

4. OTHER BOUNDS VIA DISCRETE GRÜSS INEQUALITY

Using Grüss' inequality for weighted means (see Lemma 3.3), we point out the following reverse of Jensen's inequality.

LEMMA 4.1. *Let $f,\ x_i,\ t_i$ be as in Lemma 1.4. If there exist real constants $m, M \in \mathring{I}$such that $m \leq x_i \leq M$ for all $i \in \{1, \dots, n\}$, then we have the inequality:*

$$0 \leq \frac{1}{T_n} \sum_{i=1}^{n} t_i f(x_i) - f\left(\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i\right) \leq \frac{1}{4}(M-m)(f'(M) - f'(m)), \tag{4.1}$$

where $f' : \mathring{I} \to \mathbb{R}$ is the derivative of f.

PROOF. In Grüss' inequality put $a_i := x_i$ and $b_i := f'(x_i)$. As f is convex, f' is monotonic nondecreasing and therefore $f'(m) \leq b_i \leq f'(M)$. Applying the Grüss inequality,

$$\frac{1}{T_n} \sum_{i=1}^{n} t_i x_i f'(x_i) - \frac{1}{T_n} \sum_{i=1}^{n} t_i x_i \cdot \frac{1}{T_n} \sum_{i=1}^{n} t_i f'(x_i) \leq \frac{1}{4}(M-m)(f'(M) - f'(m))$$

and by the inequality (1.9) we deduce (4.1). ∎

REMARK 4.1. Similar results can be obtained using other Grüss type inequalities.

The following reverse inequality for the Csiszár f-divergence holds [**43**].

THEOREM 4.2. *Let $f,\ p$ and q be as in Theorem 1.5. If there exist real numbers r, R such that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1, \dots, n\}$, then,*

$$0 \leq I_f(p,q) - Q_n f\left(\frac{P_n}{Q_n}\right) \leq \frac{1}{4}(R-r)(f'(R) - f'(r)) Q_n. \tag{4.2}$$

PROOF. Apply Lemma 4.1 for $f = f,\ t_i = q_i$ and $x_i = \frac{p_i}{q_i}\ (i = 1, \dots, n)$. ∎

The following particular inequalities are noted:

$$0 \leq KL(p,q) - P_n \log\left(\frac{P_n}{Q_n}\right) \leq \frac{P_n}{4}(R-r)[\log(R) - \log(r)] \leq \frac{P_n}{4} \cdot \frac{(R-r)^2}{\sqrt{Rr}}. \tag{4.3}$$

Indeed, if we choose $f(t) = t \log t$ in (4.2), we obtain the Kullback-Leibler divergence. The last inequality in (4.3) follows by the well known inequality between the *geometric mean* $G(a,b) = \sqrt{ab}$ and the *logarithmic mean* $L(a,b) := \frac{b-a}{\log b - \log a}\ \ (a, b > 0,\ a \neq b)$, i.e.

$$L(a,b) \geq G(a,b) \quad \text{for all } a, b > 0,\ a \neq b. \tag{4.4}$$

In addition, if in (4.2) we put $f(t) = -\log t$, we deduce that,

$$0 \leq KL(q,p) - Q_n \log\left(\frac{Q_n}{P_n}\right) \leq \frac{Q_n}{4} \cdot \frac{(R-r)^2}{Rr}, \tag{4.5}$$

provided that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1, \dots, n\}$.

Now, if in (4.2) we choose $f(t) = t^{\alpha}\ \ (\alpha > 1)\,,\ t > 0$, then we get the following inequality for the $\alpha-$order Renyi entropy,

$$0 \leq D_{\alpha}(p,q) - P_n^{\alpha} Q_n^{1-\alpha} \leq \frac{\alpha}{4} \cdot Q_n (R-r) \left(R^{\alpha-1} - r^{\alpha-1}\right), \tag{4.6}$$

provided that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1,\ldots,n\}$.

If we apply Theorem 4.2 for the Bhattacharyya distance, we get,

$$0 \leq \sqrt{P_n Q_n} - B(p,q) \leq \frac{1}{8} \cdot \frac{(R-r)\left(\sqrt{R}-\sqrt{r}\right)}{\sqrt{rR}} \cdot Q_n, \tag{4.7}$$

provided that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1,\ldots,n\}$.

Finally, if we apply Theorem 4.2 for the J-divergence, we can obtain the inequality,

$$0 \leq J(p,q) - (P_n - Q_n) \ln \frac{P_n}{Q_n} \leq \frac{1}{4}(R-r)\left(\ln \frac{R}{r} + \frac{R-r}{rR}\right) Q_n, \tag{4.8}$$

provided that $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1,\ldots,n\}$.

5. Further Reverse Inequalities

We start with the following result.

THEOREM 5.1 (Dragomir, 2003 [**44**]). *Let $f : [0,\infty) \to \mathbb{R}$ be a convex mapping on the interval $[r,R] \subset [0,\infty)$ with $r \leq 1 \leq R$. If $p,q \in \mathbb{P}^n$ and $r \leq \frac{p_i}{q_i} \leq R$ for all $i \in \{1,...,n\}$, then we have the inequality*

$$I_f(p,q) \leq \frac{R-1}{R-r} \cdot f(r) + \frac{1-r}{R-r} \cdot f(R). \tag{5.1}$$

PROOF. As f is convex on $[r,R]$, we may write that

$$f(tr + (1-t)R) \leq tf(r) + (1-t) f(R) \tag{5.2}$$

for all $t \in [0,1]$.

Choose $t = \frac{R-x}{R-r}$, $x \in [r,R]$. Then $1-t = \frac{x-r}{R-r}$ and from (5.2) we deduce

$$f(x) \leq \frac{R-x}{R-r} \cdot f(r) + \frac{x-r}{R-r} \cdot f(R) \tag{5.3}$$

for all $x \in [r,R]$, as a simple calculation shows that $\frac{R-x}{R-r} \cdot r + \frac{x-r}{R-r} \cdot R = x$. Put in (5.3) $x = \frac{p_i}{q_i}$, $i \in \{1,...,n\}$, to get

$$f\left(\frac{p_i}{q_i}\right) \leq \frac{R - \frac{p_i}{q_i}}{R-r} \cdot f(r) + \frac{\frac{p_i}{q_i} - r}{R-r} \cdot f(R) \tag{5.4}$$

for all $i \in \{1,...,n\}$.

If we multiply (5.4) by $q_i \geq 0$, sum over i and take into account that

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$$

then by (5.4) we obtain (5.1). ∎

The following result also holds.

THEOREM 5.2 (Dragomir, 2003 [**44**]). *Let $f : [0, \infty) \to \mathbb{R}$ be differentiable convex on $[r, R]$ and p, q be as in Theorem 5.1. Then we have the inequality:*

$$0 \leq \frac{R-1}{R-r} \cdot f(r) + \frac{1-r}{R-r} \cdot f(R) - I_f(p,q) \leq \frac{f'(R) - f'(r)}{R-r} \cdot [(R-1)(1-r) - D_{\chi^2}(p,q)] \leq \frac{1}{4}(R-r)[f'(R) - f'(r)], \tag{5.5}$$

where $D_{\chi^2}(\cdot,\cdot)$ is the chi-square divergence.

PROOF. Since the mapping f is differentiable convex, we can write

$$f(u) - f(v) \geq f'(v)(u-v) \tag{5.6}$$

for all $u, v \in (r, R)$.

Now, assume that $\alpha, \beta \geq 0$ and $\alpha + \beta > 0$. Then, by (5.6), we have

$$f\left(\frac{\alpha a + \beta b}{\alpha + \beta}\right) - f(a) \geq f'(a)\left(\frac{\alpha a + \beta b}{\alpha + \beta} - a\right) = \frac{\beta}{\alpha + \beta} \cdot f'(a)(b-a) \tag{5.7}$$

and

$$f\left(\frac{\alpha a + \beta b}{\alpha + \beta}\right) - f(b) \geq f'(b)\left(\frac{\alpha a + \beta b}{\alpha + \beta} - b\right) = -\frac{\alpha}{\alpha + \beta} \cdot f'(b)(b-a). \tag{5.8}$$

Now, if we multiply (5.7) by α and (5.8) by β and add the obtained results, we get

$$(\alpha + \beta) f\left(\frac{\alpha a + \beta b}{\alpha + \beta}\right) - \alpha f(a) - \beta f(b) \geq \frac{\alpha\beta}{\alpha + \beta}(b-a)(f'(a) - f'(b))$$

which is equivalent to:

$$0 \leq \frac{\alpha f(a) + \beta f(b)}{\alpha + \beta} - f\left(\frac{\alpha a + \beta b}{\alpha + \beta}\right) \leq \frac{\alpha\beta}{(\alpha + \beta)^2}(f'(b) - f'(a))(b-a). \tag{5.9}$$

Now, if in (5.9) we choose $\alpha = R - x$, $\beta = x - r$, $a = r$, $b = R$, then we obtain

$$0 \leq \frac{(R-x) f(r) + (x-r) f(R)}{R-r} - f(x) \leq \frac{(R-x)(x-r)}{R-r}(f'(R) - f'(r)). \tag{5.10}$$

If in (5.10), we choose $x = \frac{p_i}{q_i}$ and then multiply with q_i we get

$$\frac{(Rq_i - p_i) f(r) + (p_i - rq_i) f(R)}{R-r} - q_i f\left(\frac{p_i}{q_i}\right) \leq \frac{(Rq_i - p_i)(p_i - rq_i)}{(R-r) q_i}(f'(R) - f'(r)) \tag{5.11}$$

for all $i \in \{1, ..., n\}$.

If we sum over i in (5.11) and take into consideration that

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1,$$

we get

$$\frac{(R-1) f(r) + (1-r) f(R)}{R-r} - I_f(p,q) \leq \frac{(f'(R) - f'(r))}{R-r} \sum_{i=1}^{n} \frac{(Rq_i - p_i)(p_i - rq_i)}{q_i}. \tag{5.12}$$

However,

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} \frac{(Rq_i - p_i)(p_i - rq_i)}{q_i} \\ &= R - \sum_{i=1}^{n} \frac{p^2(y)}{q_i} - rR + r = R + r - rR - 1 - D_{\chi^2}(p,q) \\ &= (R-1)(1-r) - D_{\chi^2}(p,q). \end{aligned}$$

As

$$(R-1)(1-r) \leq \frac{1}{4}(R-r)^2 \quad \text{and} \quad D_{\chi^2}(p,q) \geq 0,$$

the last inequality is obvious. ∎

The following results also holds.

THEOREM 5.3 (Dragomir, 2003 [**44**]). *Assume that the function $f : [0,\infty) \to \mathbb{R}$ is twice differentiable on $[r,R]$ and*

$$m \leq f''(t) \leq M \quad \textit{for all } t \in [r,R]. \tag{5.13}$$

If the probability distributions $p,q \in \mathbb{P}^n$ satisfy the conditions of Theorem 5.1, then we have the inequality:

$$\begin{aligned} &\frac{1}{2} m \left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] \\ &\leq \frac{R-1}{R-r} \cdot f(r) + \frac{1-r}{R-r} \cdot f(R) - I_f(p,q) \\ &\leq \frac{1}{2} M \left[(R-1)(1-r) - D_{\chi^2}(p,q)\right]. \end{aligned} \tag{5.14}$$

PROOF. Define the function $f_m : [0,\infty) \to \mathbb{R}$, $f_m(t) = f(t) - \frac{1}{2} m t^2$. Then f_m is twice differentiable and $f_m''(t) = f''(t) - m \geq 0$, $t \in [r,R]$, which shows that f_m is convex on $[r,R]$.

If we write the inequality (5.1) for the convex mapping f_m, we obtain

$$I_{f - \frac{1}{2} m (\cdot)^2}(p,q) \leq \frac{R-1}{R-r}\left[f(r) - \frac{1}{2} m r^2\right] + \frac{1-r}{R-r}\left[f(R) - \frac{1}{2} m R^2\right]. \tag{5.15}$$

However,

$$\begin{aligned}
&I_{f-\frac{1}{2}m(\cdot)^2}(p,q) \\
&= I_f(p,q) - \frac{1}{2}m\left[\sum_{i=1}^{n}\frac{p^2(y)}{q_i} - 1 + 1\right] \\
&= I_f(p,q) - \frac{1}{2}mD_{\chi^2}(p,q) - \frac{1}{2}m
\end{aligned}$$

and then, by (5.15), we can get

$$\begin{aligned}
&\frac{R-1}{R-r}\cdot f(r) + \frac{1-r}{R-r}\cdot f(R) - I_f(p,q) \\
&\geq \frac{1}{2}mR^2\cdot\frac{(1-r)}{R-r} + \frac{1}{2}mr^2\cdot\frac{(R-1)}{R-r} - \frac{1}{2}mD_{\chi^2}(p,q) - \frac{1}{2}m
\end{aligned} \tag{5.16}$$

Nonetheless, the right hand side of (5.16) is

$$\frac{1}{2}m\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right]$$

and the first inequality in (5.14) is obtained.

The second inequality follows by a similar argument applied for the mapping $f_m(t) := \frac{1}{2}Mt^2 - f(t)$. We omit the details. ∎

COROLLARY 5.4 (Dragomir, 2003 [**44**]). *With the assumptions in Theorem 5.3, and if* $m \geq 0$*, then*

$$\begin{aligned}
0 &\leq \frac{1}{2}m\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] \\
&\leq \frac{R-1}{R-r}\cdot f(r) + \frac{1-r}{R-r}\cdot f(R) - I_f(p,q).
\end{aligned} \tag{5.17}$$

PROOF. We only have to prove the fact that

$$D_{\chi^2}(p,q) \leq (R-1)(1-r), \tag{5.18}$$

which follows by the fact that (see the proof of Theorem 5.2)

$$0 \leq \sum_{i=1}^{n}\frac{(Rq_i - p_i)(p_i - rq_i)}{q_i} = (R-1)(1-r) - D_{\chi^2}(p,q).$$

∎

6. APPLICATIONS FOR PARTICULAR DIVERGENCES

Before we point out some applications of the above results, we would like to recall the following special means:

$$L(\alpha,\beta) := \begin{cases} \beta & \text{if } \alpha = \beta; \\ \frac{\beta-\alpha}{\ln\beta - \ln\alpha} & \text{if } \beta \neq \alpha,\ \alpha,\beta > 0 \ \ (\textit{logarithmic mean}) \end{cases}$$

and

$$I(\alpha,\beta) := \begin{cases} \beta & \text{if } \alpha = \beta; \\ \frac{1}{e}\left(\frac{\beta^\beta}{\alpha^\alpha}\right)^{\frac{1}{\beta-\alpha}} & \text{if } \beta \neq \alpha,\ \ (\textit{identric mean}). \end{cases}$$

(1) *Kullback-Leibler Divergence.* Consider the convex mapping $f:(0,\infty)\to\mathbb{R}$, $f(t)=t\ln t$. Then

$$I_f(p,q)=\sum_{i=1}^{n}p_i\ln\left[\frac{p_i}{q_i}\right]=D(p,q),$$

where $D(p,q)$ is the *Kullback-Leibler distance.*

PROPOSITION 6.1. *Let $p,q\in\mathbb{P}^n$ with the property that:*

$$r\le\frac{p_i}{q_i}\le R \text{ for all } i\in\{1,...,n\}. \tag{6.1}$$

Then we have the inequality

$$D(p,q)\le\ln I(r,R)-\frac{G^2(r,R)}{L(r,R)}+1, \tag{6.2}$$

where $I(\cdot,\cdot)$ is the identric mean, $L(\cdot,\cdot)$ is the logarithmic mean and $G(\cdot,\cdot)$ is the usual geometric mean.

PROOF. We apply Theorem 5.1 for $f(t)=t\ln t$ to get

$$\begin{aligned}D(p,q)&\le\frac{R-1}{R-r}r\ln r+\frac{1-r}{R-r}R\ln R\\&=\frac{R\ln R-r\ln r}{R-r}-rR\cdot\frac{\ln R-\ln r}{R-r}\\&=\ln I(r,R)+1-\frac{G^2(r,R)}{L(r,R)}\end{aligned}$$

and the inequality (6.2) is proved. ∎

PROPOSITION 6.2. *With the assumptions of Proposition 6.1, we have*

$$\begin{aligned}0&\le\ln I(r,R)-\frac{G^2(r,R)}{L(r,R)}+1-D(p,q)\\&\le\frac{(R-1)(1-r)-D_{\chi^2}(p,q)}{L(r,R)}.\end{aligned} \tag{6.3}$$

The proof follows by Theorem 5.2 applied for $f(t)=t\ln t$, and taking into account that

$$\frac{f'(R)-f'(r)}{R-r}=\frac{1}{L(r,R)}.$$

Using Theorem 5.3, we may be able to improve the inequality (6.3) as follows.

PROPOSITION 6.3. *Let $p,q\in\mathbb{P}^n$ satisfy the condition (6.1). Then we have the inequality:*

$$\begin{aligned}&\frac{1}{2R}\left[(R-1)(1-r)-D_{\chi^2}(p,q)\right]\\&\le\ln I(r,R)-\frac{G^2(r,R)}{L(r,R)}+1-D(p,q)\\&\le\frac{1}{2r}\left[(R-1)(1-r)-D_{\chi^2}(p,q)\right].\end{aligned} \tag{6.4}$$

PROOF. We have $f''(t) = \frac{1}{t}$, $t \in [r, R]$ and then

$$\frac{1}{R} \leq f''(t) \leq \frac{1}{r}, \ t \in [r, R].$$

Applying Theorem 5.3 for $f(t) = t \ln t$, we obtain (6.4). ∎

Now, assume that $f(t) = -\ln t$, which is a convex mapping as well.
We have

$$I_f(p,q) = -\sum_{i=1}^{n} q_i \ln \left[\frac{p_i}{q_i}\right] = \sum_{i=1}^{n} q_i \ln \left[\frac{q_i}{p_i}\right] = D(q,p).$$

Using Theorem 5.1, we may state the following proposition.

PROPOSITION 6.4. *Let $p, q \in \mathbb{P}^n$ with the property that (6.1) holds. Then we have the inequality:*

$$D(q,p) \leq \ln I\left(\frac{1}{r}, \frac{1}{R}\right) - \frac{1}{L(r,R)} + 1. \tag{6.5}$$

PROOF. Applying the inequality (5.1) for $f(t) = -\ln t$, we may write that

$$\begin{aligned}
&D(q,p) \\
&\leq \frac{(R-1)(-\ln r) + (1-r)(-\ln R)}{R-r} \\
&= \frac{r\ln R - R\ln r}{R-r} - \frac{\ln R - \ln r}{R-r} = \frac{rR\left(\frac{1}{R}\ln R - \frac{1}{r}\ln r\right)}{R-r} - \frac{1}{L(r,R)} \\
&= \frac{\frac{1}{r}\ln\frac{1}{r} - \frac{1}{R}\ln\frac{1}{R}}{\frac{1}{r} - \frac{1}{R}} - \frac{1}{L(r,R)} = \ln I\left(\frac{1}{r}, \frac{1}{R}\right) + 1 - \frac{1}{L(r,R)}
\end{aligned}$$

and the inequality (6.5) is proved. ∎

PROPOSITION 6.5. *Let p, q be as in Proposition 6.1. Then*

$$\begin{aligned}
0 &\leq \ln I\left(\frac{1}{r}, \frac{1}{R}\right) - \frac{1}{L(r,R)} + 1 - D(q,p) \\
&\leq \frac{1}{G^2(r,R)}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right].
\end{aligned} \tag{6.6}$$

The proof follows by Theorem 5.2 applied for the function $f(t) = -\ln t$, and taking into account that

$$\frac{f'(R) - f'(r)}{R-r} = \frac{1}{rR} = \frac{1}{G^2(r,R)}.$$

The inequality (6.6) can be improved as follows.

PROPOSITION 6.6. *Let p, q be as in Proposition 6.1. Then*

$$\begin{aligned}
&\frac{1}{2R^2}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] \\
&\leq \ln I\left(\frac{1}{r}, \frac{1}{R}\right) - \frac{1}{L(r,R)} + 1 - D(q,p) \\
&\leq \frac{1}{2r^2}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right].
\end{aligned} \tag{6.7}$$

The proof is obvious by Theorem 5.3, taking into account that $f''(t) = \frac{1}{t^2}$ and $\frac{1}{R^2} \le f''(t) \le \frac{1}{r^2}$ for all $t \in [r, R]$.

(2) *Hellinger discrimination.* Consider the convex mapping $f : [0, \infty) \to \mathbb{R}$, $f(t) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$. Then

$$I_f(p, q) = \frac{1}{2}\sum_{i=1}^{n} q_i \left(\sqrt{\frac{p_i}{q_i}} - 1\right)^2 = \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{p_i} - \sqrt{q_i}\right)^2 = h^2(p, q),$$

where $h^2(p, q)$ is the *Hellinger discrimination.*

PROPOSITION 6.7. *With the assumptions of Proposition 6.1, we have*

$$h^2(p, q) \le \frac{\left(\sqrt{R} - 1\right)\left(1 - \sqrt{r}\right)}{\sqrt{R} + \sqrt{r}}. \tag{6.8}$$

PROOF. We apply Theorem 5.1 for $f(t) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$ to get

$$\begin{aligned}
& h^2(p, q) \\
& \le \frac{(R-1)\frac{1}{2}\left(\sqrt{r} - 1\right)^2 + (1 - r)\frac{1}{2}\left(\sqrt{R} - 1\right)^2}{R - r} \\
& = \frac{\frac{1}{2}\left(\sqrt{R} - 1\right)\left(\sqrt{r} - 1\right)}{R - r}\left[\left(\sqrt{R} + 1\right)\left(1 - \sqrt{r}\right) + \left(1 + \sqrt{r}\right)\left(\sqrt{R} - 1\right)\right] \\
& = \frac{\left(\sqrt{R} - 1\right)\left(\sqrt{r} - 1\right)\left(\sqrt{R} - \sqrt{r}\right)}{R - r} = \frac{\left(\sqrt{R} - 1\right)\left(1 - \sqrt{r}\right)}{\sqrt{R} + \sqrt{r}},
\end{aligned}$$

and the inequality (6.8) is proved. ∎

Using Theorem 5.2, we may state the following proposition as well.

PROPOSITION 6.8. *With the assumptions of Proposition 6.1, we have*

$$\begin{aligned}
0 & \le \frac{\left(\sqrt{R} - 1\right)\left(1 - \sqrt{r}\right)}{\sqrt{R} + \sqrt{r}} - h^2(p, q) \\
& \le \frac{1}{4(r - R)A\left(\sqrt{r}, \sqrt{R}\right)}\left[(R - 1)(1 - r) - D_{\chi^2}(p, q)\right],
\end{aligned} \tag{6.9}$$

where $A(\cdot, \cdot)$ is the arithmetic mean.

The proof is obvious by Theorem 5.2 applied for $f(t) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$, taking into account that $f'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}}$, and

$$\frac{f'(R) - f'(r)}{R - r} = \frac{\sqrt{R} - \sqrt{r}}{2\sqrt{rR}(R - r)} = \frac{1}{2\sqrt{rR}\left(\sqrt{R} + \sqrt{r}\right)}.$$

Finally, by the use of Theorem 5.3, we may state:

PROPOSITION 6.9. *Assume that* $p, q \in \mathbb{P}^n$ *are as in Proposition 6.1. Then*

$$\begin{aligned}
(6.10) \qquad & \frac{1}{8\sqrt{R^3}}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right] \\
& \leq \frac{\left(\sqrt{R}-1\right)(1-\sqrt{r})}{\sqrt{R}+\sqrt{r}} - h^2(p,q) \\
& \leq \frac{1}{8\sqrt{r^3}}\left[(R-1)(1-r) - D_{\chi^2}(p,q)\right].
\end{aligned}$$

The proof follows by Theorem 5.3 applied for the mapping $f(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$ for which $f''(t) = \frac{1}{4\sqrt{t^3}}$ and, obviously,

$$\frac{1}{4\sqrt{R^3}} \leq f''(t) \leq \frac{1}{4\sqrt{r^3}} \text{ for all } t \in [r, R].$$

CHAPTER 2

Jensen Type Inequalities for (m, M)-Convex Functions

The concept of (m, M)-convex functions defined on convex subsets in normed linear spaces is introduced and some inequalities of Jensen's type are derived. Applications for norm inequalities and f-divergence measure are provided as well.

1. General Results in Normed Spaces

Let $(X, \|\cdot\|)$ be a real or complex normed linear space, $C \subseteq X$ a convex subset of X and $f : C \to \mathbb{R}$, see [**37**].

Definition 1.1. Let α, $\beta \in \mathbb{R}$.

(i) The mapping f will be called $\alpha-$*lower convex* on C if f-$\frac{\alpha}{2} \cdot \|\cdot\|^2$ is a convex mapping on C;

(ii) The mapping f will be called $\beta-$*upper convex* on C if $\frac{\beta}{2} \cdot \|\cdot\|^2 - f$ is a convex mapping on C;

(iii) The mapping f will be called $(\alpha, \beta) -$ *convex* on C if it is both $\alpha-$lower convex and β-upper convex on C.

Note that if f is (α, β)-convex on C, then $\alpha \leq \beta$.

Indeed, if f is (α, β) $-$convex, then f-$\frac{\alpha}{2} \cdot \|\cdot\|^2$ and $\frac{\beta}{2} \cdot \|\cdot\|^2 - f$ are convex, which clearly implies that the sum

$$f - \frac{\alpha}{2} \cdot \|\cdot\|^2 + \frac{\beta}{2} \cdot \|\cdot\|^2 - f = \frac{\beta - \alpha}{2} \cdot \|\cdot\|^2$$

is convex, and then $\beta \geq \alpha$.

Taking into account the above, when we talk about an (α, β) $-$convex function, we can assume without loss of generality that $\alpha = m \leq M = \beta$.

The following theorem holds [**37**].

Theorem 1.1 (Dragomir, 2001 [**37**]). *Let* $f : C \subseteq X \to \mathbb{R}$, C *be convex on* X, $x_i \in C$, $p_i \geq 0$ $(i = 1, \dots, n)$ *with* $P_n := \sum_{i=1}^{n} p_i > 0$.

(i) *If* f *is* $\alpha-$*lower convex on* C, *then we have the following inequality (for* $\alpha \geq 0$ *- refinement of Jensen's inequality),*

$$\frac{\alpha}{2} \cdot \frac{1}{P_n^2} \left[P_n \sum_{i=1}^{n} p_i \|x_i\|^2 - \left\| \sum_{i=1}^{n} p_i x_i \right\|^2 \right] \leq \frac{1}{P_n} \sum_{i=1}^{n} p_i f(x_i) - f\left(\frac{1}{P_n} \sum_{i=1}^{n} p_i x_i \right). \tag{1.1}$$

(ii) *If f is $\beta-$upper convex on C, then we have the following inequality (which is a counterpart of Jensen's inequality if f is convex),*

$$\frac{1}{P_n}\sum_{i=1}^{n}p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right) \leq \frac{\beta}{2}\cdot\frac{1}{P_n^2}\left[P_n\sum_{i=1}^{n}p_i\|x_i\|^2 - \left\|\sum_{i=1}^{n}p_i x_i\right\|^2\right]. \tag{1.2}$$

(iii) *If f is $(m, M)-$convex on C, then we have the following 'sandwich' inequality,*

$$\begin{aligned}&\frac{m}{2}\cdot\frac{1}{P_n^2}\left[P_n\sum_{i=1}^{n}p_i\|x_i\|^2 - \left\|\sum_{i=1}^{n}p_i x_i\right\|^2\right]\\ &\leq \frac{1}{P_n}\sum_{i=1}^{n}p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right)\\ &\leq \frac{M}{2}\cdot\frac{1}{P_n^2}\left[P_n\sum_{i=1}^{n}p_i\|x_i\|^2 - \left\|\sum_{i=1}^{n}p_i x_i\right\|^2\right].\end{aligned} \tag{1.3}$$

PROOF. If $g : C \subseteq X \to \mathbb{R}$ is a convex mapping on C, $x_i \in C$, $p_i \geq 0$ with $P_n > 0$, then, Jensen's inequality holds

$$g\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right) \leq \frac{1}{P_n}\sum_{i=1}^{n}p_i g(x_i). \tag{1.4}$$

(i) Let $g(x) = f(x) - \frac{\alpha}{2}\cdot\|x\|^2$, then g is convex on C and by (1.4) we get,

$$f\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right) - \frac{\alpha}{2}\left\|\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right\|^2 \leq \frac{1}{P_n}\sum_{i=1}^{n}p_i\left[f(x_i) - \frac{\alpha}{2}\|x_i\|^2\right],$$

which is clearly equivalent to (1.1).

(ii) Let $h(x) = \frac{\beta}{2}\cdot\|x\|^2 - f(x)$, then h is convex on C, and by (1.4), we obtain (1.2).

(iii) Follows by (i) and (ii).

∎

The following corollary for inner product spaces holds.

COROLLARY 1.2 (Dragomir, 2001 [**37**]). *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space, $\|\cdot\| := \sqrt{\langle\cdot,\cdot\rangle}$, $C \subseteq X$ a convex subset on X, $f : C \to \mathbb{R}$, and $x_i \in C$, $p_i \geq 0$ $(i = 1, \ldots, n)$ with $P_n > 0$.*

(i) *If f is $\alpha-$lower convex on C, then,*

$$\begin{aligned}&\frac{\alpha}{2P_n^2}\sum_{1\leq i<j\leq n}p_i p_j\|x_i - x_j\|^2\\ &\leq \frac{1}{P_n}\sum_{i=1}^{n}p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right).\end{aligned} \tag{1.5}$$

(ii) *If f is $\beta-$upper convex on C, then,*

$$\frac{1}{P_n}\sum_{i=1}^{n}p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right) \le \frac{\beta}{2P_n^2}\sum_{1\le i<j\le n}p_i p_j\|x_i - x_j\|^2 . \tag{1.6}$$

(iii) *If f is $(m,M)-$convex on C, then,*

$$\begin{aligned}\frac{m}{2}\cdot\frac{1}{P_n^2}\sum_{1\le i<j\le n}p_i p_j\|x_i - x_j\|^2 &\le \frac{1}{P_n}\sum_{i=1}^{n}p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n}p_i x_i\right)\\ &\le \frac{M}{2}\cdot\frac{1}{P_n^2}\sum_{1\le i<j\le n}p_i p_j\|x_i - x_j\|^2 .\end{aligned} \tag{1.7}$$

PROOF. The argument follows by Theorem 1.1, taking into account that, for inner products, we have:

$$\begin{aligned}&\sum_{1\le i<j\le n}p_i p_j\|x_i - x_j\|^2\\ &= \frac{1}{2}\sum_{i,j=1}^{n}p_i p_j\|x_i - x_j\|^2 = \frac{1}{2}\sum_{i,j=1}^{n}p_i p_j\left[\|x_i\|^2 - 2\operatorname{Re}\langle x_i, x_j\rangle + \|x_j\|^2\right]\\ &= \frac{1}{2}\left[\sum_{i=1}^{n}p_i\|x_i\|^2 P_n - 2\operatorname{Re}\left\langle\sum_{i=1}^{n}p_i x_i, \sum_{j=1}^{n}p_j x_j\right\rangle + P_n\sum_{j=1}^{n}p_j\|x_j\|^2\right]\\ &= P_n\sum_{i=1}^{n}p_i\|x_i\|^2 - \left\|\sum_{i=1}^{n}p_i x_i\right\|^2 .\end{aligned}$$

∎

REMARK 1.1. Results for the case of mappings defined on real intervals have been obtained by Andrica and Raşa in [**2**].

Furthermore, assume that,

$$\Delta(x) := \max_{1\le i<j\le n}\|x_i - x_j\|$$

and

$$\delta(x) := \min_{1\le i<j\le n}\|x_i - x_j\| .$$

The following corollary also holds.

COROLLARY 1.3 (Dragomir, 2001 [**37**]). *Let X, C, f, x_i, p_i $(i = 1, \ldots, n)$ be as in Corollary 1.2.*

(i) *If f is $\alpha-$lower convex on C with $\alpha > 0$, then we have the following refinement of Jensen's inequality:*

$$0 < \frac{\alpha}{4}\left(1 - \frac{\sum_{i=1}^{n} p_i^2}{P_n^2}\right)\delta^2(x) \leq \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right). \tag{1.8}$$

(ii) *If f is convex and $\beta-$upper convex on C, then,*

$$0 \leq \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \leq \frac{\beta}{4}\left(1 - \frac{\sum_{i=1}^{n} p_i^2}{P_n^2}\right)\Delta^2(x). \tag{1.9}$$

(iii) *If f is $(m, M)-$convex on C with $m > 0$, then,*

$$\begin{aligned} &\frac{m}{4}\left(1 - \frac{\sum_{i=1}^{n} p_i^2}{P_n^2}\right)\delta^2(x) \\ &\leq \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) \\ &\leq \frac{M}{4}\left(1 - \frac{\sum_{i=1}^{n} p_i^2}{P_n^2}\right)\Delta^2(x). \end{aligned} \tag{1.10}$$

To prove the above corollary we use the following lemma which is also of inherent interest.

LEMMA 1.4 (Dragomir, 2001 [**37**]). *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space and $x_i \in X$, $p_i \geq 0$ $(i = 1, \ldots, n)$ with $P_n > 0$, then,*

$$\frac{1}{2}\cdot\left(1 - \frac{\sum_{i=1}^{n} p_i^2}{P_n^2}\right)\delta^2(x) \leq \frac{1}{P_n}\sum_{i=1}^{n} p_i\|x_i\|^2 - \left\|\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right\|^2 \leq \frac{1}{2}\cdot\left(1 - \frac{\sum_{i=1}^{n} p_i^2}{P_n^2}\right)\Delta^2(x). \tag{1.11}$$

PROOF. As above, we have,

$$\frac{1}{P_n}\sum_{i=1}^{n} p_i\|x_i\|^2 - \left\|\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right\|^2 = \frac{1}{2P_n^2}\sum_{i,j=1}^{n} p_i p_j\|x_i - x_j\|^2 = \frac{1}{P_n^2}\sum_{1\leq i<j\leq n} p_i p_j\|x_i - x_j\|^2. \tag{1.12}$$

Obviously,

$$\delta^2(x)\frac{1}{P_n^2}\sum_{1\leq i<j\leq n} p_i p_j \leq \frac{1}{P_n^2}\sum_{1\leq i<j\leq n} p_i p_j\|x_i - x_j\|^2 \leq \Delta^2(x)\frac{1}{P_n^2}\sum_{1\leq i<j\leq n} p_i p_j. \tag{1.13}$$

On the other hand,

$$\sum_{i,j=1} p_i p_j = \left(\sum_{i=1}^{n} p_i \right)^2 = P_n^2$$

and

$$\sum_{i,j=1}^{n} p_i p_j = \sum_{1 \le i < j \le n} p_i p_j + \sum_{1 \le j < i \le n} p_i p_j + \sum_{i=j=1}^{n} p_i p_j$$
$$= 2 \sum_{1 \le i < j \le n} p_i p_j + \sum_{i=1}^{n} p_i^2,$$

from which we obtain,

$$\sum_{1 \le i < j \le n} p_i p_j = \frac{1}{2} \left(P_n^2 - \sum_{i=1}^{n} p_i^2 \right). \tag{1.14}$$

Now, using (1.12) - (1.14), we get the desired inequality (1.11). ∎

2. Applications for f-Divergence

In this section we apply some of the above results for f-divergence and obtain other inequalities that are similar, in a sense, to those presented above [**37**].

THEOREM 2.1 (Dragomir, 2001 [**37**]). *Let $f : \mathbb{R}_+ \to \mathbb{R}$ and $p, q \in \mathbb{R}_+^n$ with $P_n = Q_n > 0$.*

(i) If f is $\alpha-$lower convex on $\mathbb{R}_+$, then,

$$\frac{\alpha}{2} \cdot D_{\chi^2}(p, q) \le I_f(p, q) - Q_n f(1). \tag{2.1}$$

(ii) If f is $\beta-$upper convex on $\mathbb{R}_+$, then,

$$I_f(p, q) - Q_n f(1) \le \frac{\beta}{2} \cdot D_{\chi^2}(p, q). \tag{2.2}$$

(iii) If f is $(m, M)-$convex on $\mathbb{R}_+$, then,

$$\frac{m}{2} \cdot D_{\chi^2}(p, q) \le I_f(p, q) - Q_n f(1) \le \frac{M}{2} \cdot D_{\chi^2}(p, q), \tag{2.3}$$

where $D_{\chi^2}(p, q)$ is the χ^2-distance, i.e.,

$$D_{\chi^2}(p, q) = \sum_{i=1}^{n} \frac{p_i^2 - q_i^2}{q_i}.$$

PROOF. We follow the proof in [**37**].

We use Theorem 1.1 in which $f = f$, $C = \mathbb{R}_+$, $X = \mathbb{R}_+$, $p_i \to q_i$, $x_i = \frac{p_i}{q_i}$. We have,

$$\frac{1}{P_n} \sum_{i=1}^{n} p_i f(x_i) - f\left(\frac{1}{P_n} \sum_{i=1}^{n} p_i x_i \right)$$
$$= \frac{1}{Q_n} \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i} \right) - f\left(\frac{1}{Q_n} \sum_{i=1}^{n} q_i \cdot \frac{p_i}{q_i} \right) = \frac{1}{Q_n} I_f(p, q) - f(1)$$

and

$$\begin{aligned}
&\frac{1}{P_n^2}\left[P_n\sum_{i=1}^{n}p_ix_i^2-\left(\sum_{i=1}^{n}p_ix_i\right)^2\right]\\
&=\frac{1}{Q_n^2}\left[Q_n\sum_{i=1}^{n}q_i\left(\frac{p_i}{q_i}\right)^2-\left(\sum_{i=1}^{n}q_i\cdot\frac{p_i}{q_i}\right)^2\right]=\frac{1}{Q_n^2}\left[Q_n\sum_{i=1}^{n}\frac{p_i^2}{q_i}-P_n^2\right]\\
&=\frac{1}{Q_n}D_{\chi^2}(p,q).
\end{aligned}$$

Now, by (1.1)-(1.3), we obtain the desired results (2.1)-(2.3). ∎

COROLLARY 2.2 (Dragomir, 2001 [**37**]). *Let $f:\mathbb{R}_+\to\mathbb{R}$ be normalised, i.e., $f(1)=0$, and $p,q\in\mathbb{R}_+^n$ with $P_n=Q_n>0$.*

(i) *If f is $\alpha-$lower convex on $\mathbb{R}_+$, then,*

$$\frac{\alpha}{2}\cdot D_{\chi^2}(p,q)\le I_f(p,q). \tag{2.4}$$

(ii) *If f is β-upper convex on $\mathbb{R}_+$, then,*

$$I_f(p,q)\le\frac{\beta}{2}\cdot D_{\chi^2}(p,q). \tag{2.5}$$

(iii) *If f is $(m,M)-$convex on $\mathbb{R}_+$, then,*

$$\frac{m}{2}\cdot D_{\chi^2}(p,q)\le I_f(p,q)\le\frac{M}{2}\cdot D_{\chi^2}(p,q). \tag{2.6}$$

In practical applications, it is important to have sufficient conditions for the mapping f so that it will be $\alpha-$lower convex, $\beta-$upper convex, or (m,M)-convex on a certain interval I of $\mathbb{R}_+$.

PROPOSITION 2.3. *Let $f:I\subseteq\mathbb{R}_+\to\mathbb{R}$ be a differentiable mapping on $\mathring{I}$($\mathring{I}$is the interior of I).*

(i) *If there exists $\alpha\in\mathbb{R}$ such that,*

$$\frac{f'(x_2)-f'(x_1)}{x_2-x_1}\ge\alpha \text{ for all } x_2>x_1,\ x_2,\ x_1\in\mathring{I}; \tag{2.7}$$

then f is $\alpha-$lower convex on I.

(ii) *If there exists $\beta\in\mathbb{R}$ such that,*

$$\frac{f'(x_2)-f'(x_1)}{x_2-x_1}\le\beta \text{ for all } x_2>x_1,\ x_2,\ x_1\in\mathring{I}; \tag{2.8}$$

then f is $\beta-$upper convex on I.

(iii) *If there exist $m,M\in\mathbb{R}$ such that $m<M$ and*

$$m\le\frac{f'(x_2)-f'(x_1)}{x_2-x_1}\le M \text{ for all } x_2>x_1,\ x_2,\ x_1\in\mathring{I}, \tag{2.9}$$

then f is $(m,M)-$convex on I.

PROOF. It is well-known that a differentiable mapping $g:I\to\mathbb{R}$ is convex iff g' is monotonic nondecreasing on $\mathring{I}$, i.e., $g'(x_2)\ge g'(x_1)$ for all $x_2>x_1,x_1,x_2\in\mathring{I}$.

Applying this criterion to the mapping $f(x)-\frac{\alpha}{2}\cdot x^2$, $\frac{\beta}{2}\cdot x^2-f(x)$, $f(x)-m\cdot\frac{x^2}{2}$, $M\cdot\frac{x^2}{2}-f(x)$, we obtain the desired results. ∎

PROPOSITION 2.4. *Let* $f : I \subseteq \mathbb{R}_+ \to \mathbb{R}$ *be a twice differentiable mapping on* $\mathring{I}$*.*

(i) *If there exists* $\alpha \in \mathbb{R}$ *such that,*

$$f'' (x) \geq \alpha \ \text{ for all } x \in \mathring{I}, \tag{2.10}$$

then f *is* $\alpha-$*lower convex on* I*.*

(ii) *If there exists* $\beta \in \mathbb{R}$ *such that,*

$$f'' (x) \leq \beta \ \text{ for all } x \in \mathring{I}, \tag{2.11}$$

then f *is* $\beta-$*upper convex on* I*.*

(iii) *If there exist* $m, M \in \mathbb{R}$ *such that* $m < M$ *and*

$$m \leq f'' (x) \leq M \ \text{ for all } x \in \mathring{I}, \tag{2.12}$$

then f *is* $(m, M)-$*convex on* I*.*

The proof is obvious by the well-known fact that a twice differentiable mapping $g : I \to \mathbb{R}$ is convex iff $g'' (x) \geq 0$ for all $x \in \mathring{I}$.

We omit the details.

In what follows, we apply the previous result to some well-known information measures which are Csiszár f-divergences for some appropriate choices of the mapping f.

3. APPLICATIONS FOR SOME PARTICULAR f-DIVERGENCES

PROPOSITION 3.1. *Let* $p, q \in \mathbb{R}^n_+$ *with* $P_n = Q_n > 0$*. Denote* $r_i := \frac{p_i}{q_i}$*,* $(i = 1, \ldots, n)$*.*

(i) *If* $r_i \leq R$*,* $i = 1, \ldots, n$*, then,*

$$KL (q, p) \geq \frac{1}{2R^2} D_{\chi^2} (p, q) . \tag{3.1}$$

(ii) *If* $r_i \geq r > 0$*, then,*

$$\frac{1}{2r^2} D_{\chi^2} (p, q) \geq KL (q, p) . \tag{3.2}$$

(iii) *If* $0 < m \leq r_i \leq M < \infty$*, then,*

$$\frac{1}{2M^2} D_{\chi^2} (p, q) \leq KL (q, p) \leq \frac{1}{2m^2} D_{\chi^2} (p, q) . \tag{3.3}$$

PROOF. As $f (t) = -\log t$, then $f' (t) = -\frac{1}{t}$, $f'' (t) = \frac{1}{t^2}$ and if $t \in [a, b] \subset (0, \infty)$, then $\frac{1}{b^2} \leq f'' (t) \leq \frac{1}{a^2}$, $t \in [a, b]$. Using Proposition 2.4 and Corollary 2.2, we deduce the desired results. ∎

We know that for $f (t) = t \log t$,

$$I_f (p, q) = KL (p, q) .$$

If we apply Corollary 2.2 for the mapping $f (t) = t \log t$, we can, therefore, state the following proposition.

PROPOSITION 3.2. *Let* $p, q \in \mathbb{R}^n_+$ *with* $P_n = Q_n > 0$ *and* $r_i := \frac{p_i}{q_i}$ $(i = 1, \ldots, n)$*.*

(i) *If* $r_i \leq R$*,* $(i = 1, \ldots, n)$*, then,*

$$\frac{1}{2R} D_{\chi^2} (p, q) \leq KL (p, q) . \tag{3.4}$$

(ii) *If* $r_i \geq r > 0$*, then,*

$$KL (p, q) \leq \frac{1}{2r} D_{\chi^2} (p, q) . \tag{3.5}$$

(iii) *If* $0 < m \le r_i \le M$, $(i = 1, \dots, n)$, *then,*

$$\frac{1}{2M} D_{\chi^2}(p,q) \le KL(p,q) \le \frac{1}{2m} D_{\chi^2}(p,q). \tag{3.6}$$

Now, let us consider the Rényi $\alpha-$distance ($\alpha > 1$).

PROPOSITION 3.3. *Let* $p, q \in \mathbb{R}^n_+$, $P_n = Q_n > 0$ *and* $r_i := \frac{p_i}{q_i}$, $(i = 1, \dots, n)$.

(i) *We have the inequalities,*

$$D_\alpha(p,q) - P_n \ge \begin{cases} \frac{\alpha(\alpha-1)r^{\alpha-2}}{2} D_{\chi^2}(p,q) \\ \textit{if } \alpha \in [2,\infty) \textit{ and } r_i \ge r > 0, \ (i = 1, \dots, n), \\ \frac{\alpha(\alpha-1)R^{\alpha-2}}{2} D_{\chi^2}(p,q) \\ \textit{if } \alpha \in (1,2) \textit{ and } r_i \le R \ \ (i = 1, \dots, n). \end{cases} \tag{3.7}$$

(ii)

$$D_\alpha(p,q) - P_n \le \begin{cases} \frac{\alpha(\alpha-1)}{2} R^{\alpha-2} D_{\chi^2}(p,q) \\ \textit{if } \alpha \in [2,\infty) \textit{ and } r_i \le R \ (i = 1, \dots, n), \\ \frac{\alpha(\alpha-1)}{2} r^{\alpha-2} D_{\chi^2}(p,q) \\ \textit{if } \alpha \in (1,2) \textit{ and } r_i \ge r \ (i = 1, \dots, n). \end{cases} \tag{3.8}$$

(iii) *If* $m \le r_i \le M$ *and* $\alpha \in [2,\infty)$, *then,*

$$\begin{aligned} \frac{\alpha(\alpha-1)}{2} m^{\alpha-2} D_{\chi^2}(p,q) &\le D_\alpha(p,q) - P_n \\ &\le \frac{\alpha(\alpha-1)}{2} M^{\alpha-2} D_{\chi^2}(p,q). \end{aligned} \tag{3.9}$$

If $\alpha \in (1,2)$, *then the reverse inequality holds in (3.9).*

Now consider *Hellinger discrimination* [**9**].

PROPOSITION 3.4. *Let* $p, q \in \mathbb{R}^n_+$ *with* $P_n = Q_n > 0$. *Denote* $r_i := \frac{p_i}{q_i}$ $(i = 1, \dots, n)$.

(i) *If* $r_i \le R$ $(i = 1, \dots, n)$, *then,*

$$\frac{1}{8\sqrt{R^3}} D_{\chi^2}(p,q) \le h^2(p,q). \tag{3.10}$$

(ii) *If* $r_i \ge r > 0$, *then,*

$$h^2(p,q) \le \frac{1}{8\sqrt{r^3}} D_{\chi^2}(p,q). \tag{3.11}$$

(iii) *If* $0 < m \le r_i \le M$, *then,*

$$\frac{1}{8\sqrt{M^3}} D_{\chi^2}(p,q) \le h^2(p,q) \le \frac{1}{8\sqrt{m^3}} D_{\chi^2}(p,q). \tag{3.12}$$

PROOF. As $f(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$, we have $f'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}}$, $f''(t) = \frac{1}{4\sqrt{t^3}}$ and then

$$\alpha = \inf_{0<t\leq R} f''(t) = \frac{1}{4\sqrt{R^3}},$$
$$\beta = \sup_{t\geq r} f''(t) = \frac{1}{4\sqrt{r^3}}.$$

Using Corollary 2.2 for f as above, we obtain the desired inequalities. ∎

Consider now the Bhattacharyya distance.

PROPOSITION 3.5. *Let* $p, q \in \mathbb{R}_+^n$ *with* $P_n = Q_n > 0$. *Denote* $r_i := \frac{p_i}{q_i}$ $(i = 1, \ldots, n)$.

(i) *If* $r_i \leq R$ $(i = 1, \ldots, n)$, *then,*

$$\frac{1}{8\sqrt{R^3}} D_{\chi^2}(p,q) \leq P_n - B(p,q). \tag{3.13}$$

(ii) *If* $r_i \geq r > 0$ $(i = 1, \ldots, n)$, *then,*

$$P_n - B(p,q) \leq \frac{1}{8\sqrt{r^3}} D_{\chi^2}(p,q). \tag{3.14}$$

(iii) *If* $0 < m \leq r_i \leq M$, *then,*

$$\frac{1}{8\sqrt{M^3}} D_{\chi^2}(p,q) \leq P_n - B(p,q) \leq \frac{1}{8\sqrt{m^3}} D_{\chi^2}(p,q). \tag{3.15}$$

The proof follows by Theorem 2.1 applied for the mapping $f(t) = -\sqrt{t}$.
Now consider the *Harmonic distance* (see for example [**88**])

$$M(p,q) := \sum_{i=1}^{n} \frac{2p_i q_i}{p_i + q_i}.$$

If $f(t) = -\frac{2t}{t+1}$, $t \in (0, \infty)$, then obviously,

$$f'(t) = -\frac{2}{(t+1)^2}, \quad t > 0$$
$$f''(t) = \frac{4}{(t+1)^3}, \quad t > 0$$

and

$$I_f(p,q) = -M(p,q).$$

PROPOSITION 3.6. *Let* $p, q \in \mathbb{R}_+^n$ *with* $P_n = Q_n > 0$. *Denote* $r_i := \frac{p_i}{q_i}$ $(i = 1, \ldots, n)$.

(i) *If* $r_i \leq R$ $(i = 1, \ldots, n)$, *then,*

$$\frac{2}{(R+1)^3} D_{\chi^2}(p,q) \leq P_n - M(p,q). \tag{3.16}$$

(ii) *If* $r_i \geq r > 0$, *then,*

$$P_n - M(p,q) \leq \frac{2}{(r+1)^3} D_{\chi^2}(p,q). \tag{3.17}$$

(iii) *If* $0 < m \leq r_i \leq M$, *then we have the 'sandwich' inequality,*

$$\frac{2}{(M+1)^3} D_{\chi^2}(p,q) \leq P_n - M(p,q) \leq \frac{2}{(m+1)^3} D_{\chi^2}(p,q). \tag{3.18}$$

The proof follows by Theorem 2.1 applied for the mapping $f(t) = -\frac{2t}{t+1}$.

Finally, consider *Jeffreys' distance* (see for example [**88**]).

PROPOSITION 3.7. *Let $p, q \in \mathbb{R}_+^n$ with $P_n = Q_n > 0$. Denote $r_i := \frac{p_i}{q_i}$ $(i = 1, \dots, n)$.*

(i) *If $r_i \leq R$ $(i = 1, \dots, n)$, then,*

$$\frac{R+1}{2R^2} D_{\chi^2}(p, q) \leq J(p, q). \tag{3.19}$$

(ii) *If $r_i \geq r > 0$ $(i = 1, \dots, n)$, then,*

$$J(p, q) \leq \frac{r+1}{2r^2} D_{\chi^2}(p, q). \tag{3.20}$$

(iii) *If $0 < m \leq r_i \leq M$, then,*

$$\frac{M+1}{2M^2} D_{\chi^2}(p, q) \leq J(p, q) \leq \frac{m+1}{2m^2} D_{\chi^2}(p, q). \tag{3.21}$$

CHAPTER 3

Inequalities in Terms of Kullback-Leibler Distance

In this chapter various inequalities for general f-divergence in terms of the well known Kullback-Leibler distance are established. Particular inequalities of interest for various other divergence measures in terms of this distance are provided.

1. Upper and Lower Bounds in the General Case

The following result concerning an upper and a lower bound for the f-divergence in terms of the Kullback-Leibler distance $KL(p,q)$ holds. This result complements, in a sense, the results presented above [**42**].

THEOREM 1.1 (Dragomir, 2003 [**42**]). *Assume that the generating mapping $f:(0,\infty)\to\mathbb{R}$ is normalised, i.e., $f(1)=0$ and satisfies the assumptions,*

(i) *f is twice differentiable on (r,R), where $0\le r\le 1\le R\le\infty$;*

(ii) *there exist real constants m,M such that*

$$m\le tf''(t)\le M \quad \text{for all } t\in(r,R). \tag{1.1}$$

If p,q are discrete probability distributions satisfying the assumption,

$$r\le r_i:=\frac{p_i}{q_i}\le R \quad \text{for all } i\in\{1,\dots,n\}, \tag{1.2}$$

then we have the inequality,

$$mKL(p,q)\le I_f(p,q)\le MKL(p,q). \tag{1.3}$$

PROOF. Define the mapping $F_m:(0,\infty)\to\mathbb{R}$, $F_m(t)=f(t)-mt\ln t$, then $F_m(\cdot)$ is normalised, twice differentiable and since,

$$F_m''(t)=f''(t)-\frac{m}{t}=\frac{1}{t}(tf''(t)-m)\ge 0 \tag{1.4}$$

for all $t\in(r,R)$, it follows that $F_m(\cdot)$ is convex on (r,R). Applying the nonnegativity property of the f-divergence functional for $F_m(\cdot)$ and the linearity property, we have,

$$\begin{aligned} 0\le I_{F_m}(p,q)&=I_f(p,q)-mI_{(\cdot)\ln(\cdot)}(p,q)\\ &=I_f(p,q)-mKL(p,q) \end{aligned} \tag{1.5}$$

from which the first inequality in (1.3) is obtained.

Define $F_M:(0,\infty)\to\mathbb{R}$, $F_M(t):=Mt\ln t-f(t)$, which is obviously normalised, twice differentiable and by (1.1), convex on (r,R). Applying the nonnegativity property of f-divergence for F_M, we obtain the second part of (1.3). ∎

REMARK 1.1. If in (1.1) we have the strict inequality for any $t\in(r,R)$, then the mappings F_m and F_M are strictly convex and the case of equality holds in (1.3) iff $p=q$.

REMARK 1.2. It is important to note that if f is twice differentiable on $(0,\infty)$ and $0<m\le tf''(t)\le M<\infty$ for any $t\in(0,\infty)$, then inequality (1.3) holds for any probability distributions p,q.

The following theorem concerning the convexity property of the f-divergence also holds [**42**].

THEOREM 1.2 (Dragomir, 2003 [**42**]). *Assume that f satisfies the assumptions (i) and (ii) from Theorem 1.1. If $p^{(j)}$, $q^{(j)}$ $(j=1,2)$ are probability distributions satisfying (1.2), i.e.,*

$$r \le \frac{p_i^{(j)}}{q_i^{(j)}} \le R \quad \textit{for all } i \in \{1,\dots,n\} \textit{ and } j \in \{1,2\}, \tag{1.6}$$

then,

$$r \le \frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} \le R \quad \textit{for all } i \in \{1,\dots,n\} \textit{ and } \lambda \in [0,1] \tag{1.7}$$

and

$$\begin{aligned} m\Big[&KL\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &\quad - \lambda KL\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) KL\left(p^{(2)}, q^{(2)}\right)\Big] \\ &\le I_f\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &\qquad - \lambda I_f\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) I_f\left(p^{(2)}, q^{(2)}\right) \\ &\le M\Big[KL\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &\qquad - \lambda KL\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) KL\left(p^{(2)}, q^{(2)}\right)\Big] \end{aligned} \tag{1.8}$$

for all $\lambda \in [0,1]$.

PROOF. We follow the proof in [**42**].

By (1.6),

$$r\lambda q_i^{(1)} \le \lambda p_i^{(1)} \le \lambda R q_i^{(1)} \quad \text{for all } i \in \{1,\dots,n\} \tag{1.9}$$

and

$$r(1-\lambda) q_i^{(2)} \le (1-\lambda) p_i^{(2)} \le R(1-\lambda) q_i^{(2)} \quad \text{for all } i \in \{1,\dots,n\}. \tag{1.10}$$

Summing (1.9) and (1.10), we obtain (1.7).

It is already known that the mappings F_m, F_M as defined in Theorem 1.1 are convex and normalised.

Applying the "Joint Convexity Principle" for $I_{F_m}(\cdot,\cdot)$, i.e.,

$$\begin{aligned} I_{F_m}&\left(\lambda\left(p^{(1)}, q^{(1)}\right) + (1-\lambda)\left(p^{(2)}, q^{(2)}\right)\right) \\ &\le \lambda I_{F_m}\left(p^{(1)}, q^{(1)}\right) + (1-\lambda) I_{F_m}\left(p^{(2)}, q^{(2)}\right) \end{aligned} \tag{1.11}$$

and rearranging the terms, we obtain the first inequality in (1.8).

The second inequality follows likewise if we apply the same property to the f-divergence $I_{F_M}(\cdot,\cdot)$. ∎

REMARK 1.3. If $m > 0$ in (1.1), then the inequality (1.3) is a better result than the positivity property of the f-divergence. The same will apply for the joint convexity of the f-divergence if $m > 0$.

Using inequality (1.4) which holds for f differentiable convex and normalised for p, q probability distributions, we can state the following theorem as well [**42**].

THEOREM 1.3 (Dragomir, 2003 [**42**]). *Let $f : [0, \infty) \to \mathbb{R}$ be a normalised mapping, i.e., $f(1) = 0$ and which satisfies the assumptions:*

(i) *f is twice differentiable on (r, R), where $0 \le r \le 1 \le R \le \infty$;*

(ii) *there exist constants m, M such that,*

$$m \le t f''(t) \le M \quad \text{for all } t \in (r, R). \tag{1.12}$$

If p, q are discrete probability distributions satisfying the assumption

$$r \le r_i = \frac{p_i}{q_i} \le R \quad \text{for all } i \in \{1, \dots, n\}, \tag{1.13}$$

then we have the inequality,

$$\begin{aligned} I_{f'}\left(\frac{p^2}{q}, p\right) - I_{f'}(p, q) - MD(q, p) &\\ \le I_f(p, q) &\\ \le I_{f'}\left(\frac{p^2}{q}, p\right) - I_{f'}(p, q) - mD(q, p). \end{aligned} \tag{1.14}$$

PROOF. We follow the proof in [**42**].

We know (see the proof of Theorem 1.1) that the mapping $F_m : (0, \infty) \to \mathbb{R}$, $F_m(t) = f(t) - mt \ln t$ is normalised, twice differentiable and convex on (r, R).

If we apply the second inequality from (1.4) for F_m, we may write:

$$I_{F_m}(p, q) \le I_{F'_m}\left(\frac{p^2}{q}, p\right) - I_{F'_m}(p, q). \tag{1.15}$$

However,

$$\begin{aligned} I_{F_m}(p, q) &= I_f(p, q) - mKL(q, p), \\ I_{F'_m}\left(\frac{p^2}{q}, p\right) &= I_{f'(\cdot) - m[\ln(\cdot)+1]}\left(\frac{p^2}{q}, p\right) \\ &= I_{f'}\left(\frac{p^2}{q}, p\right) - mI_{\ln(\cdot)}\left(\frac{p^2}{q}, p\right) - m \\ &= I_{f'}\left(\frac{p^2}{q}, p\right) + mKL\left(p, \frac{p^2}{q}\right) - m \end{aligned}$$

and

$$I_{F'_m}(p, q) = I_{f'}(p, q) + mKL(q, p) - m.$$

Consequently, by (1.15), we have,

$$\begin{aligned} &I_f(p, q) - mKL(p, q) \\ &\le I_{f'}\left(\frac{p^2}{q}, p\right) + mKL\left(p, \frac{p^2}{q}\right) - m - I_{f'}(p, q) - mKL(q, p) + m \\ &= I_{f'}\left(\frac{p^2}{q}, p\right) + m\left(KL\left(p, \frac{p^2}{q}\right) - KL(q, p)\right) - I_{f'}(p, q). \end{aligned}$$

As a simple computation shows that $KL\left(p, \frac{p^2}{q}\right) = -KL(p, q)$, the second inequality in (1.14) is proved.

Consider $F_M(t) := Mt\ln t - f(t)$, which is obviously normalised, twice differentiable and convex on (r,R).

If we apply the second inequality from (1.4) for F_M, we may write:

$$I_{F_M}(p,q) \le I_{F_M'}\left(\frac{p^2}{q},p\right) - I_{F_M'}(p,q). \tag{1.16}$$

However,

$$\begin{aligned} I_{F_M}(p,q) &= MKL(p,q) - I_f(p,q); \\ I_{F_M'}\left(\frac{p^2}{q},p\right) &= -MKL\left(p,\frac{p^2}{q}\right) + M - I_{f'}\left(\frac{p^2}{q},p\right); \\ I_{F_M'}(p,q) &= -MKL(q,p) + M - I_{f'}(p,q) \end{aligned}$$

and then, by (1.16), we get,

$$\begin{aligned} &MKL(p,q) - I_f(p,q) \\ &\qquad \le -MKL\left(p,\frac{p^2}{q}\right) + M - I_{f'}\left(\frac{p^2}{q},p\right) + MKL(q,p) - M + I_{f'}(p,q), \end{aligned}$$

which is equivalent to the first part of (1.14). ∎

REMARK 1.4. The inequality (1.14) is obviously equivalent to,

$$mKL(q,p) \le I_{f'}\left(\frac{p^2}{q},p\right) - I_{f'}(p,q) - I_f(p,q) \le MKL(q,p).$$

The above results have natural applications when the Kullback-Leibler distance is compared with a number of other divergence measures arising in Information Theory.

2. SOME PARTICULAR CASES

Using Theorem 1.1, we are able to point out the following particular cases which may be of interest in Information Theory.

PROPOSITION 2.1. *Let p,q be two probability distributions with the property that*

$$0 < r \le \frac{p_i}{q_i} = r_i \le R < \infty \quad \textit{for all } i \in \{1,\dots,n\}, \tag{2.1}$$

then,

$$\frac{1}{R}KL(p,q) \le KL(q,p) \le \frac{1}{r}KL(p,q). \tag{2.2}$$

PROOF. Consider the mapping $f : [r,R] \to \mathbb{R}$, $f(t) = -\ln t$. Define $g(t) = tf''(t) = t\cdot\left(\frac{1}{t^2}\right) = \frac{1}{t}$. Obviously,

$$\sup_{t\in[r,R]} g(t) = \frac{1}{r} \quad \text{and} \quad \inf_{t\in[r,R]} g(t) = \frac{1}{R}.$$

Also,

$$\begin{aligned} I_f(p,q) &= -\sum_{i=1}^n q_i \ln\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^n q_i \ln\left(\frac{q_i}{p_i}\right) \\ &= KL(q,p). \end{aligned}$$

Using (1.3) with $m = \frac{1}{R}$ and $M = \frac{1}{r}$, we deduce the desired inequality. ∎

COROLLARY 2.2. *With the above assumptions for p and q, we have:*

$$r \leq \frac{KL(p,q)}{KL(q,p)} \leq R. \tag{2.3}$$

COROLLARY 2.3. *Assume that p, q satisfy the condition,*

$$\left| \frac{p_i}{q_i} - 1 \right| \leq \varepsilon \quad \text{for all } i \in \{1, \ldots, n\}, \tag{2.4}$$

then,

$$\left| \frac{KL(p,q)}{KL(q,p)} - 1 \right| \leq \varepsilon.$$

The following proposition connecting the χ^2-distance with the Kullback-Leibler distance also holds.

PROPOSITION 2.4. *Let p, q be two probability distributions satisfying the condition (2.1), then we have the inequality:*

$$2r \leq \frac{D_{\chi^2}(p,q)}{KL(p,q)} \leq 2R. \tag{2.5}$$

PROOF. Consider the mapping $f : [r, R] \to \mathbb{R}$, $f(t) = (t-1)^2$. Define $g(t) = tf''(t) = 2t$, then, obviously,

$$\sup_{t \in [r,R]} g(t) = 2R \quad \text{and} \quad \inf_{t \in [r,R]} g(t) = 2r.$$

Since

$$I_f(p,q) = D_{\chi^2}(p,q),$$

then, applying (1.3) for $m = 2r$ and $M = 2R$, we deduce the desired inequality. ∎

REMARK 2.1. The following inequality is well known in the literature

$$KL(p,q) \leq D_{\chi^2}(p,q). \tag{2.6}$$

For a simple proof of this fact as well as for different applications in Information Theory, see [**11**].
Now, observe that from the first inequality in (2.5), we have,

$$KL(p,q) \leq \frac{1}{2r} D_{\chi^2}(p,q). \tag{2.7}$$

We note that if $\frac{1}{2r} \leq 1$ i.e., $r \geq \frac{1}{2}$, the inequality (2.7) is better than (2.6).

The following corollary is obvious.

COROLLARY 2.5. *Assume that the probability distributions p, q satisfy the condition (2.4), then,*

$$\frac{1}{2} \left| \frac{D_{\chi^2}(p,q)}{KL(p,q)} - 2 \right| \leq \varepsilon. \tag{2.8}$$

PROPOSITION 2.6. *Assume that the probability distributions p, q satisfy the condition (2.1), then we have,*

$$\frac{1}{4\sqrt{R}} KL(p,q) \leq h^2(p,q) \leq \frac{1}{4\sqrt{r}} KL(p,q). \tag{2.9}$$

PROOF. Consider the mapping $f(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$, giving $f'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}}$ and $f''(t) = \frac{1}{4\sqrt{t^3}}$. Define $g : [r, R] \to \mathbb{R}$ where g is given by,

$$g(t) = tf''(t) = \frac{1}{4\sqrt{t}}.$$

Clearly,

$$\sup_{t\in[r,R]} g(t) = \frac{1}{4\sqrt{r}} \text{ and } \inf_{t\in[r,R]} g(t) = \frac{1}{4\sqrt{R}}.$$

Since

$$I_f(p,q) = h^2(p,q),$$

then by (1.3) for $m = \frac{1}{4\sqrt{R}}$ and $M = \frac{1}{4\sqrt{r}}$, we deduce the desired inequality (2.9). ∎

REMARK 2.2. The following inequality is well known in the literature (see for example [**27**]):

$$KL(p,q) \geq 2h^2(p,q) \tag{2.10}$$

for two p, q probability distributions.

From the second inequality in (2.9), we have,

$$KL(p,q) \geq 4\sqrt{r}h^2(p,q). \tag{2.11}$$

We remark that if $4\sqrt{r} \geq 2$, i.e., $r \geq \frac{1}{4}$, then the inequality in (2.11) is better than (2.10).

The following result establishes a connection between the triangular discrimination Δ (see Remark 3.1) and the Kullback-Leibler distance.

PROPOSITION 2.7. *Assume that the probability distributions p, q satisfy the condition (2.1).*

(i) *If $0 < r \leq \frac{1}{2}$, then we have,*

$$8\min\left\{\frac{r}{(r+1)^3}, \frac{R}{(R+1)^3}\right\} KL(p,q) \leq \Delta(p,q) \leq \frac{32}{27}KL(p,q). \tag{2.12}$$

(ii) *If $\frac{1}{2} < r < 1$, then,*

$$\frac{8R}{(R+1)^3}KL(p,q) \leq \Delta(p,q) \leq \frac{8r}{(r+1)^3}KL(p,q). \tag{2.13}$$

PROOF. Consider the mapping $f(t) = \frac{(t-1)^2}{t+1}$. We have,

$$f'(t) = 1 - \frac{4}{(t+1)^2}$$

and

$$f''(t) = \frac{8}{(t+1)^3}.$$

Define

$$g : [r, R] \to \mathbb{R}, \quad g(t) = tf''(t) = \frac{8t}{(t+1)^3}, \quad t \in [r, R],$$

giving,

$$g'(t) = \frac{8(1-2t)}{(t+1)^4},$$

which shows that g has its maximum realized at $t_0 = \frac{1}{2}$ and

$$\max_{t\in(0,\infty)} g(t) = g\left(\frac{1}{2}\right) = \frac{32}{27}.$$

We have the two cases:

1) If $0 < r \leq \frac{1}{2}$, then,

$$\sup_{t\in[r,R]} g(t) = \frac{32}{27} \quad \text{and}$$

$$\inf_{t\in[r,R]} g(t) = \min\left[g(r), g(R)\right] = \min\left\{\frac{8r}{(r+1)^3}, \frac{8R}{(R+1)^3}\right\}.$$

2) If $\frac{1}{2} < r < 1$, then,

$$\sup_{t\in[r,R]} g(t) = g(r) = \frac{8r}{(r+1)^3} \quad \text{and}$$

$$\inf_{t\in[r,R]} g(t) = g(R) = \frac{8R}{(R+1)^3}.$$

Applying (1.3), we deduce (2.12) and (2.13). We omit the details. ∎

REMARK 2.3. It is clear, by the above arguments, that for every probability distribution we have the inequality

$$\Delta(p,q) \leq \frac{32}{27} KL(p,q). \tag{2.14}$$

We know that (see Topsoe [**127**])

$$2h^2(p,q) \leq \Delta(p,q) \leq 4h^2(p,q). \tag{2.15}$$

Now, as $KL(p,q) \geq 2h^2(p,q)$, then we obtain,

$$\Delta(p,q) \leq 2KL(p,q), \tag{2.16}$$

which is not as good as our result (2.14).

Let us now compare the *Rényi* divergence with the Kullback-Leibler distance.

PROPOSITION 2.8. *Assume that the probability distributions p, q satisfy the condition (2.1), then,*

$$\begin{aligned} \alpha(\alpha-1) r^{\alpha-1} KL(p,q) + 1 &\leq \exp\left[\alpha(\alpha-1) R_\alpha(p,q)\right] \\ &\leq \alpha(\alpha-1) R^{\alpha-1} KL(p,q) + 1 \end{aligned} \tag{2.17}$$

for $\alpha > 1$.

PROOF. Consider the mapping $f : (0,\infty) \to \mathbb{R}$, $f(t) = t^\alpha - 1$, $\alpha > 1$, giving $f'(t) = \alpha t^{\alpha-1}$ and $f''(t) = \alpha(\alpha-1) t^{\alpha-2}$. Define $g : [r,R] \to \mathbb{R}$, $g(t) = tf''(t) = \alpha(\alpha-1) t^{\alpha-1}$. It is obvious that,

$$\sup_{t\in[r,R]} g(t) = \alpha(\alpha-1) R^{\alpha-1} \quad \text{and} \quad \inf_{t\in[r,R]} g(t) = \alpha(\alpha-1) r^{\alpha-1}.$$

Now, observe that $f(1) = 0$, i.e., f is normalised and so we can apply the inequality (1.3) getting,

$$\alpha(\alpha-1) r^{\alpha-1} KL(p,q) \leq \sum_{i=1}^{n} q_i \left[\left(\frac{p_i}{q_i}\right)^\alpha - 1\right] \leq \alpha(\alpha-1) R^{\alpha-1} KL(p,q),$$

i.e.,

$$\alpha(\alpha-1)\,r^{\alpha-1}KL(p,q)+1\le\rho_{\alpha}(p,q)\le\alpha(\alpha-1)\,R^{\alpha-1}KL(p,q)+1$$

and the proposition is proved. ∎

We define the *Bhattacharyya distance* by (see **[11]**) $\gamma(p,q)=-\ln[B(p,q)]$.

PROPOSITION 2.9. *Assume that the probability distributions* p,q *satisfy the condition (2.1), then,*

$$4\sqrt{r}\,[1-\exp[-\gamma(p,q)]]\le KL(p,q)\le 4\sqrt{R}\,[1-\exp[-\gamma(p,q)]]. \tag{2.18}$$

PROOF. Consider the mapping $f:(0,\infty)\to\mathbb{R}$, $f(t)=\sqrt{t}-1$, then f is normalised, $f'(t)=\frac{1}{2}t^{-\frac{1}{2}}$, $f''(t)=-\frac{1}{4}t^{-\frac{3}{2}}$. Define $g:[r,R]\to\mathbb{R}$, $g(t)=tf''(t)=-\frac{1}{4}t^{-\frac{1}{2}}$. It is obvious that,

$$\sup_{t\in[r,R]}g(t)=g(R)=-\frac{1}{4\sqrt{R}},\quad \inf_{t\in[r,R]}g(t)=g(r)=-\frac{1}{4\sqrt{r}}.$$

Applying (1.3), we have:

$$-\frac{1}{4\sqrt{r}}KL(p,q)\le\sum_{i=1}^{n}q_i\left(\sqrt{\frac{p_i}{q_i}}-1\right)\le-\frac{1}{4\sqrt{R}}KL(p,q),$$

i.e.,

$$1-\frac{1}{4\sqrt{r}}KL(p,q)\le B(p,q)\le 1-\frac{1}{4\sqrt{R}}KL(p,q),$$

which is equivalent to (2.18). ∎

We define the *harmonic* divergence by $m(p,q):=1-M(p,q)$, where,

$$M(p,q):=\sum_{i=1}^{n}\frac{2p_iq_i}{p_i+q_i}.$$

PROPOSITION 2.10. *Assume that* p,q *are two discrete probability distributions, then,*

$$0\le m(p,q)\le\frac{16}{27}KL(p,q). \tag{2.19}$$

PROOF. Consider the mapping $f:(0,\infty)\to\mathbb{R}$, $f(t)=\frac{2t}{t+1}-1$, then f is normalised and

$$f'(t)=\frac{2}{(t+1)^2},\quad f''(t)=\frac{-4}{(t+1)^3}.$$

Define $g:[r,R]\to\mathbb{R}$, $g(t)=tf''(t)=\frac{-4t}{(t+1)^3}$, then,

$$g'(t)=\frac{4(2t-1)}{(t+1)^4}.$$

It is clear that g is monotonic decreasing on $\left[0,\frac{1}{2}\right)$ and monotonic increasing on $\left(\frac{1}{2},\infty\right)$. We have,

$$\inf_{t\in(0,\infty)}g(t)=g\left(\frac{1}{2}\right)=-\frac{16}{27},$$
$$\sup_{t\in(0,\infty)}g(t)=0.$$

Applying the inequality (1.3) for $m=-\frac{16}{27}$ and $M=0$, we deduce,

$$-\frac{16}{27}KL(p,q)\le\sum_{i=1}^{n}q_i\left\{\left[\frac{2\frac{p_i}{q_i}}{\frac{p_i}{q_i}+1}\right]-1\right\}\le 0,$$

which is equivalent to,

$$-\frac{16}{27}KL(p,q) \leq M(p,q) - 1 \leq 0$$

and the inequality (2.19) is proved. ∎

The above result can be improved if we know more information about $r_i := \frac{p_i}{q_i}, i = 1, \ldots, n$.

PROPOSITION 2.11. *Assume that p, q satisfy the condition (1.2).*

(i) *If $r \in \left(0, \frac{1}{2}\right)$, then,*

$$1 - \frac{16}{27}KL(p,q) \leq M(p,q) \leq 1 - 4\min\left\{\frac{r}{(r+1)^3}, \frac{R}{(R+1)^3}\right\} KL(p,q). \tag{2.20}$$

(ii) *If $r \in \left[\frac{1}{2}, 1\right)$, then,*

$$1 - \frac{4r}{(r+1)^3}KL(p,q) \leq M(p,q) \leq 1 - \frac{4R}{(R+1)^3}KL(p,q). \tag{2.21}$$

PROOF. (1)

(i) If $r \in \left(0, \frac{1}{2}\right)$, then,

$$\begin{aligned}
-\frac{16}{27} \leq g(t) &\leq \max\{g(r), g(R)\} \\
&= \max\left\{-\frac{4r}{(r+1)^3}, -\frac{4R}{(R+1)^3}\right\} \\
&= -4\min\left\{\frac{r}{(r+1)^3}, \frac{R}{(R+1)^3}\right\}, \quad t \in [r, R].
\end{aligned}$$

and then, applying (1.3), we may write,

$$-\frac{16}{27}KL(p,q) \leq M(p,q) - 1 \leq -4\min\left\{\frac{r}{(r+1)^3}, \frac{R}{(R+1)^3}\right\} KL(p,q),$$

and the inequality (2.20) is proved.

(ii) If $r \in \left[\frac{1}{2}, 1\right)$, then,

$$g(r) \leq g(t) \leq g(R) \quad \text{for all } t \in [r, R],$$

that is,

$$-\frac{4r}{(r+1)^3} \leq g(t) \leq -\frac{4R}{(R+1)^3}, \quad t \in [r, R].$$

Applying (1.3), we deduce (2.21).

∎

Let us now consider *J-divergence* [**84**].

PROPOSITION 2.12. *Assuming that p, q satisfy the condition (1.2), then,*

$$\frac{R+1}{R}KL(p,q) \leq J(p,q) \leq \frac{r+1}{r}KL(p,q). \tag{2.22}$$

PROOF. Consider $f(t) = (t-1)\ln t$, then $f'(t) = \ln t - \frac{1}{t} + 1$ and $f''(t) = \frac{t+1}{t^2}$. Define $g(t) = tf''(t) = 1 + \frac{1}{t}$. Obviously,

$$\sup_{t\in[r,R]} g(t) = 1 + \frac{1}{r}, \quad \inf_{t\in[r,R]} g(t) = 1 + \frac{1}{R}.$$

Now, using (1.3), for $M = \frac{r+1}{1}$, $m = \frac{R+1}{R}$, we obtain the desired result. ∎

REMARK 2.4. Similar results can be obtained by applying Theorem 1.3.

CHAPTER 4

Inequalities in Terms of Hellinger Discrimination

In this chapter various inequalities for general f-divergence in terms of the well known Hellinger discrimination are established. Particular inequalities of interest for various other divergence measures in terms of this discrimination are provided.

1. General Bounds in Terms of Hellinger Discrimination

The following result concerning an upper and a lower bound for the Csiszár f-divergence in terms of the Hellinger discrimination $h^2(p,q)$ holds. These results will complement, in a sense, the ones presented above [**41**].

THEOREM 1.1 (Dragomir, 2002 [**41**]). *Assume that the generating mapping* $f:(0,\infty)\to\mathbb{R}$ *is normalized, i.e.,* $f(1)=0$ *and satisfies the assumptions,*

(i) f *is twice differentiable on* (r,R), *where* $0\le r\le 1\le R\le\infty$,

(ii) *there exists real constants* m,M *such that,*

$$m\le t^{\frac{3}{2}}f''(t)\le M\quad \textit{for all } t\in(r,R). \tag{1.1}$$

If p,q *are discrete probability distributions verifying the assumption,*

$$r\le r_i:=\frac{p_i}{q_i}\le R\quad \textit{for all } i\in\{1,\dots,n\}, \tag{1.2}$$

then we have the inequality,

$$4mh^2(p,q)\le I_f(p,q)\le 4Mh^2(p,q). \tag{1.3}$$

PROOF. We follow the proof in [**41**].

Define the mapping $H_m:(0,\infty)\to\mathbb{R}$, $H_m(t)=f(t)-2m\left(\sqrt{t}-1\right)^2$. It follows that $H_m(\cdot)$ is normalised, twice differentiable and since,

$$H_m''(t)=f''(t)-\frac{m}{t^{\frac{3}{2}}}=\frac{1}{t^{\frac{3}{2}}}\left(t^{\frac{3}{2}}f''(t)-m\right)\ge 0 \tag{1.4}$$

for all $t\in(a,b)$, this is implied by the first inequality in (1.1). Thus, the mapping $H_m(\cdot)$ is convex on (r,R).

Applying the nonnegativity property of the f-divergence functional for $H_m(\cdot)$ and the linearity, we have that,

$$\begin{aligned} 0\le I_{H_m}(p,q)&=I_f(p,q)-2mI_{\left(\sqrt{\cdot}-1\right)^2}(p,q) \\ &=I_f(p,q)-4mh^2(p,q), \end{aligned} \tag{1.5}$$

giving the first inequality in (1.3).

Define $H_M:(0,\infty)\to\mathbb{R}$, $H_M(t)=2M\left(\sqrt{t}-1\right)^2-f(t)$ which obviously is normalised, twice differentiable and, by (1.1), convex on (r,R).

Applying the nonnegativity property of f-divergence for I_{H_M}, we obtain the second part of (1.3). ∎

The following theorem concerning the convexity property of the f-divergence also holds [**41**].

THEOREM 1.2 (Dragomir, 2002 [**41**]). *Assume that f satisfies the assumptions (i) and (ii) from Theorem 1.1. If $p^{(j)}, q^{(j)}$ $(j = 1, 2)$ are probability distributions satisfying (1.2), that is,*

$$r \leq \frac{p_i^{(j)}}{q_i^{(j)}} \leq R \quad \textit{for all } i \in \{1, \ldots, n\} \textit{ and } j \in \{1, 2\}, \tag{1.6}$$

then,

$$r \leq \frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} \leq R \quad \textit{for all } i \in \{1, \ldots, n\} \textit{ and } \lambda \in [0,1] \tag{1.7}$$

and

$$\begin{aligned} 4m\big[&h^2\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &- \lambda h^2\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) h^2\left(p^{(2)}, q^{(2)}\right)\big] \\ \leq\, &I_f\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &- \lambda I_f\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) I_f\left(p^{(2)}, q^{(2)}\right) \\ \leq\, &4M\big[h^2\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &- \lambda h^2\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) h^2\left(p^{(2)}, q^{(2)}\right)\big] \end{aligned} \tag{1.8}$$

for all $\lambda \in [0,1]$.

PROOF. We follow the proof in [**41**].

By (1.6), we have

$$r\lambda q_i^{(1)} \leq \lambda p_i^{(1)} \leq \lambda R q_i^{(1)} \quad \text{for all } i \in \{1, \ldots, n\} \tag{1.9}$$

and

$$r(1-\lambda) q_i^{(2)} \leq (1-\lambda) p_i^{(2)} \leq R(1-\lambda) q_i^{(2)} \quad \text{for all } i \in \{1, \ldots, n\}. \tag{1.10}$$

Summing (1.9) and (1.10), we obtain (1.7).

It is known that the mappings H_m, H_M as defined in Theorem 1.1 are convex and normalised.

Applying the "Joint Convexity Principle" for $I_{H_m}(\cdot,\cdot)$, i.e.,

$$\begin{aligned} I_{H_m}&\left(\lambda\left(p^{(1)}, q^{(1)}\right) + (1-\lambda)\left(p^{(2)}, q^{(2)}\right)\right) \\ &\leq \lambda I_{H_m}\left(p^{(1)}, q^{(1)}\right) + (1-\lambda) I_{H_m}\left(p^{(2)}, q^{(2)}\right) \end{aligned} \tag{1.11}$$

and rearranging the terms, we obtain the first inequality in (1.8).

The second inequality follows likewise if we apply the same property to the f-divergence $I_{H_M}(\cdot,\cdot)$. ∎

REMARK 1.1. If $m > 0$ in (1.1), then the inequality (1.3) is a better result than the positivity property of the f-divergence. The same will apply for the joint convexity of the f-divergence if $m > 0$.

Using the inequality (1.4) which holds for f being a differentiable convex and normalised function, for p, q probability distributions, we can state the following theorem.

THEOREM 1.3 (Dragomir, 2002 [**41**]). *Let $f : [0, \infty) \to \mathbb{R}$ be a normalised mapping satisfying*

(i) *f is twice differentiable on (r,R), where $0\le r\le 1\le R\le\infty$;*
(ii) *there exist constants m,M such that,*

$$m\le t^{\frac{3}{2}}f''(t)\le M\quad \textit{for all } t\in(r,R)\,. \tag{1.12}$$

If p,q are discrete probability distributions verifying the assumption,

$$r\le r_i:=\frac{p_i}{q_i}\le R\quad \textit{for all } i\in\{1,\dots,n\}\,, \tag{1.13}$$

then we have the inequality,

$$\begin{aligned} &I_{f'}\left(\frac{p^2}{q},p\right)-I_{f'}(p,q)-2MC(p,q)+4Mh^2(p,q)\\ &\le I_f(p,q)\\ &\le I_{f'}\left(\frac{p^2}{q},p\right)-I_{f'}(p,q)-2mC(p,q)+4mh^2(p,q)\,, \end{aligned} \tag{1.14}$$

where $C(p,q):=\sum_{i=1}^n(q_i-p_i)\sqrt{\frac{q_i}{p_i}}$.

PROOF. We follow the proof in [**41**].

We know (see the proof of Theorem 1.1), that the mapping $H_m:[0,\infty)\to\mathbb{R}$, $H_m(t):=f(t)-2m\left(\sqrt{t}-1\right)^2$ is normalised, twice differentiable and convex on (r,R).

If we apply the second inequality from (1.4) for H_m, we may write,

$$I_{H_m}(p,q)\le I_{H'_m}\left(\frac{p^2}{q},p\right)-I_{H'_m}(p,q)\,. \tag{1.15}$$

However,

$$\begin{aligned} I_{H_m}(p,q)&=I_f(p,q)-4mh^2(p,q)\,,\\ I_{H'_m}\left(\frac{p^2}{q},p\right)&=I_{f'(\cdot)-4m\left(\frac{1}{2}-\frac{1}{2\sqrt{\cdot}}\right)}\left(\frac{p^2}{q},p\right)\\ &=I_{f'}\left(\frac{p^2}{q},p\right)-2m+2mI_{\frac{1}{\sqrt{\cdot}}}\left(\frac{p^2}{q},p\right)\\ &=I_{f'}\left(\frac{p^2}{q},p\right)-2m+2m\sum_{i=1}^n p_i\left(\frac{1}{\sqrt{\frac{p_i^2}{q_i}\cdot\frac{1}{p_i}}}\right)\\ &=I_{f'}\left(\frac{p^2}{q},p\right)-2m+2m\sum_{i=1}^n p_i\sqrt{\frac{q_i}{p_i}}\\ &=I_{f'}\left(\frac{p^2}{q},p\right)-2m+2m\sum_{i=1}^n\sqrt{p_iq_i} \end{aligned}$$

and

$$I_{H'_m}(p,q) = I_{f'}(p,q) - 2m + 2mI_{\frac{1}{\sqrt{\cdot}}}(p,q)$$
$$= I_{f'}(p,q) - 2m + 2m\sum_{i=1}^{n} q_i \left(\frac{1}{\sqrt{\frac{p_i}{q_i}}}\right)$$
$$= I_{f'}(p,q) - 2m + 2m\sum_{i=1}^{n} q_i \sqrt{\frac{q_i}{p_i}}$$

and so, by (1.15), we obtain,

$$I_f(p,q) - 4mh^2(p,q)$$
$$\leq I_{f'}\left(\frac{p^2}{q}, p\right) - 2m + 2m\sum_{i=1}^{n} p_i\sqrt{\frac{q_i}{p_i}} - I_{f'}(p,q) + 2m - 2m\sum_{i=1}^{n} q_i\sqrt{\frac{q_i}{p_i}},$$

which is equivalent to the second inequality in (1.14).

If we consider $H_M(t) := 2M\left(\sqrt{t}-1\right)^2 - f(t)$, $t \geq 0$, then we observe that $H_M(\cdot)$ is normalised, twice differentiable and convex on (r,R). Applying the second inequality from (1.4), we deduce the first part of (1.14). ∎

The above results have natural applications when the Hellinger distance is compared with a number of other divergence measures.

2. Some Particular Cases

Using Theorem 1.1, we are able to point out the following particular cases which are of interest.

PROPOSITION 2.1. *Let p, q be two probability distributions with the property that,*

$$0 < r \leq \frac{p_i}{q_i} =: q_i \leq R < \infty \quad \textit{for all } i \in \{1,\dots,n\}, \tag{2.1}$$

then,

$$4\sqrt{r}h^2(p,q) \leq KL(p,q) \leq 4\sqrt{R}h^2(p,q). \tag{2.2}$$

PROOF. Consider the mapping $f : (0,\infty) \to \mathbb{R}$, $f(t) = t\ln t$, then,

$$f''(t) = \frac{1}{t}, \quad t \in (0,\infty).$$

Consider the mapping $g : [r,R] \to \mathbb{R}$, $g(t) = t^{\frac{3}{2}} \cdot \frac{1}{t} = t^{\frac{1}{2}}$, then

$$\inf_{t\in[r,R]} g(t) = \sqrt{r}, \qquad \sup_{t\in[r,R]} g(t) = \sqrt{R}.$$

Therefore, applying (1.3) with $m = \sqrt{r}$, $M = \sqrt{R}$, we obtain (2.2). ∎

REMARK 2.1. The following inequality is well known in the literature (see for example Dacunha-Castelle [**27**]):

$$KL(p,q) \geq 2h^2(p,q) \tag{2.3}$$

for any p, q probability distributions.
From the first inequality in (2.2) we have,

$$KL(p,q) \geq 4\sqrt{r}h^2(p,q). \tag{2.4}$$

We note that if $4\sqrt{r} \geq 2$, i.e., $r \geq \frac{1}{4}$, then the inequality (2.4) is better than (2.3).

PROPOSITION 2.2. *Let p, q be two probability distributions with the property (2.1), then,*

$$\frac{4}{\sqrt{R}} h^2(p,q) \leq KL(q,p) \leq \frac{4}{\sqrt{r}} h^2(p,q). \tag{2.5}$$

PROOF. Consider the mapping $f : [r, R] \to \mathbb{R}$, $f(t) = -\ln t$. Define $g(t) = t^{\frac{3}{2}} f''(t) = \frac{1}{\sqrt{t}}$, then,

$$\sup_{t\in[r,R]} g(t) = \frac{1}{\sqrt{r}}, \quad \inf_{t\in[r,R]} g(t) = \frac{1}{\sqrt{R}}.$$

In addition,

$$I_f(p,q) = -\sum_{i=1}^{n} q_i \ln\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{n} q_i \ln\left(\frac{q_i}{p_i}\right) = D(q,p).$$

Using (2.2), we get the desired inequality (2.5). ∎

The following result for the χ^2-distance also holds.

PROPOSITION 2.3. *Let p, q be two probability distributions satisfying the condition (2.1), then,*

$$8r^{\frac{3}{2}} h^2(p,q) \leq D_{\chi^2}(p,q) \leq 8R^{\frac{3}{2}} h^2(p,q). \tag{2.6}$$

PROOF. Consider the mapping $f : (0,\infty) \to \mathbb{R}$, $f(t) = (t-1)^2$. Define $g : [r, R] \to \mathbb{R}$, $g(t) = t^{\frac{3}{2}} f''(t) = 2t^{\frac{3}{2}}$. Obviously,

$$\sup_{t\in[r,R]} g(t) = 2R^{\frac{3}{2}} \quad \text{and} \quad \inf_{t\in[r,R]} g(t) = 2r^{\frac{3}{2}}.$$

Since

$$I_f(p,q) = D_{\chi^2}(p,q),$$

then, applying the inequality (1.3) with $m = 2r^{\frac{3}{2}}$, $M = 2R^{\frac{3}{2}}$, we get the desired inequality (2.6). ∎

Now, let us consider the $J-$*divergence* [**84**].

PROPOSITION 2.4. *Let p, q be two probability distributions, then we have the inequality,*

$$8h^2(p,q) \leq J(p,q). \tag{2.7}$$

PROOF. Consider the mapping $f : (0,\infty) \to \mathbb{R}$, $f(t) = (t-1)\ln t$. Define $g : [r, R] \to \mathbb{R}$,

$$g(t) = t^{\frac{3}{2}} f''(t) = t^{\frac{1}{2}} + \frac{1}{t^{\frac{1}{2}}} \geq 2,$$

which shows that

$$\inf_{t\in(0,\infty)} g(t) = 2.$$

Since

$$I_f(p,q) = J(p,q),$$

then, applying (1.3) with $m = 2$, we get the desired inequality. ∎

If we know more about $r_i := \frac{p_i}{q_i}$ $(i = 1, \ldots, n)$, i.e., the condition (2.1) holds, then we can obtain an upper bound for $J(\cdot,\cdot)$ as follows:

PROPOSITION 2.5. *If* $0 < r \le r_i \le R < \infty$ *for all* $i \in \{1, \ldots, n\}$, *then we have,*

$$J(p,q) \le 4 \max\left\{\sqrt{r} + \frac{1}{\sqrt{r}}, \sqrt{R} + \frac{1}{\sqrt{R}}\right\} h^2(p,q). \tag{2.8}$$

PROOF. As above, we have,

$$g(t) = t^{\frac{1}{2}} + \frac{1}{t^{\frac{1}{2}}}.$$

For the mapping $h(u) = u + \frac{1}{u}$, we have,

$$h'(u) = \frac{u^2 - 1}{u^2},$$

which shows that the mapping is strictly decreasing on $(0,1)$ and strictly increasing on $(1,\infty)$. It follows that,

$$\sup_{t \in [r,R]} g(t) = \max[g(r), g(R)] = \max\left\{\sqrt{r} + \frac{1}{\sqrt{r}}, \sqrt{R} + \frac{1}{\sqrt{R}}\right\}.$$

Applying Theorem 1.1 we deduce the desired result. ∎

REMARK 2.2. Observing that

$$\sqrt{R} + \frac{1}{\sqrt{R}} - \sqrt{r} - \frac{1}{\sqrt{r}} = \frac{\left(\sqrt{R} - \sqrt{r}\right)\left(\sqrt{rR} - 1\right)}{\sqrt{rR}},$$

(2.8) can be rewritten in the equivalent form,

$$J(p,q) \le 4h^2(p,q) \times \begin{cases} \sqrt{R} + \frac{1}{\sqrt{R}} & \text{if } \ R \ge \frac{1}{r} \\ \\ \sqrt{r} + \frac{1}{\sqrt{r}} & \text{if } \ 1 \le R < \frac{1}{r} \end{cases}. \tag{2.9}$$

PROPOSITION 2.6. *Let* p, q *be two probability distributions, then, with* $M(p,g)$ *being the harmonic distance,*

$$0 \le 1 - M(p,q) \le \frac{1}{2} h^2(p,q). \tag{2.10}$$

PROOF. Consider the function $f : (0,\infty) \to \mathbb{R}$, $f(t) = 1 - \frac{2t}{t+1}$, then,

$$f'(t) = -\frac{2}{(1+t)^2}, \quad f''(t) = \frac{4}{(t+1)^3}.$$

Define the mapping

$$g(t) = t^{\frac{3}{2}} f''(t) = \frac{4t^{\frac{3}{2}}}{(t+1)^3}$$

and a simple calculation shows that,

$$g'(t) = \frac{6\sqrt{t}(1-t)}{(t+1)^4}.$$

Consequently, the mapping g is increasing on the interval $(0,1)$ and decreasing on $(1,\infty)$. Moreover,

$$\sup_{t \in (0,\infty)} = g(1) = \frac{1}{2} \text{ and } I_f(p,q) = 1 - M(p,q).$$

Applying the inequality (1.3) for $M = \frac{1}{2}$, we deduce (2.10). ∎

If we know that the condition (2.1) holds, then we can improve the first inequality in (2.10) as follows.

PROPOSITION 2.7. *Assuming that the probability distributions* p, q *satisfy (2.1), then, we have the inequality,*

$$16\min\left\{\frac{r^{\frac{3}{2}}}{(r+1)^3}, \frac{R^{\frac{3}{2}}}{(R+1)^3}\right\} h^2(p,q) \leq 1 - M(p,q). \tag{2.11}$$

PROOF. Taking into account that the mapping $g(t) = \frac{4t^{\frac{3}{2}}}{(t+1)^3}$ is monotonic increasing on $(0,1)$ and decreasing on $(1,\infty)$, we may assert that,

$$\inf_{t\in[r,R]} g(t) = \min\{g(r), g(R)\} = 4\min\left\{\frac{r^{\frac{3}{2}}}{(r+1)^3}, \frac{R^{\frac{3}{2}}}{(R+1)^3}\right\}.$$

Using (1.3), we deduce the desired lower bound (2.11). ∎

REMARK 2.3. Similar results can be stated by Applying Theorem 1.3.

CHAPTER 5

Inequalities in Terms of Variation Distance

In this chapter various inequalities for general f-divergence in terms of the well known variational distance are established. Particular inequalities of interest for various other divergence measures in terms of this distance are provided.

1. General Results

Define the generalised $r-$*variational distance* by

$$V_r(p,q) := \sum_{i=1}^{n} q_i^{1-r} |p_i - q_i|^r, \tag{1.1}$$

where p, q are probability distributions and $r \in (0,1]$. Note that for $r = 1$, we recapture the usual variational distance (or l_1-distance).

The following theorem holds [**40**].

THEOREM 1.1 (Dragomir, 2002 [**40**]). *Let $f : [0,\infty) \to \mathbb{R}$ be a normalised convex mapping of the $r - H-$Hölder type on $[r,R]$, i.e.,*

$$|f(x) - f(y)| \le H|x-y|^r \quad \textit{for all } x,y \in [r,R]. \tag{1.2}$$

It follows that,

$$0 \le I_f(p,q) \le HV_r(p,q). \tag{1.3}$$

PROOF. The proof follows that given in [**40**].

We choose in (1.2) $x = \frac{p_i}{q_i}$, $y = 1$ $(i = 1,\dots,n)$ to get,

$$\left| f\left(\frac{p_i}{q_i}\right) - f(1) \right| \le H \left| \frac{p_i}{q_i} - 1 \right|^r, \tag{1.4}$$

for all $i \in \{1,\dots,n\}$.

If we multiply (1.4) by q_i, sum the obtained inequalities and use the generalised triangle inequality, we obtain,

$$\begin{aligned} 0 \le I_f(p,q) &= \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) = \left| \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^{n} q_i f(1) \right| \\ &\le H \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - 1 \right|^r = HV_r(p,q) \end{aligned}$$

and (1.3) is proved. ∎

REMARK 1.1. If we assume that f is convex, normalised and $L-$Lipschitzian on $[r,R]$, i.e., $r = 1$ and $H = L$, then we have the inequality,

$$0 \le I_f(p,q) \le LV_r(p,q), \tag{1.5}$$

where $V(p,q)$ is the usual variational distance.

A practical result is embodied in the following corollary.

COROLLARY 1.2 (Dragomir, 2002 [**40**]). *If the mapping* $f : [r, R] \to \mathbb{R}$ *is convex, normalised, absolutely continuous on* $[r, R]$ *and* $f' \in L_\infty [r, R]$, *i.e.,* $\|f'\|_\infty := ess \sup_{t \in [r,R]} |f'(t)| < \infty$, *then we have the inequality,*

$$0 \le I_f(p, q) \le \|f'\|_\infty V_r(p, q). \tag{1.6}$$

The following theorem holds [**40**].

THEOREM 1.3 (Dragomir, 2002 [**40**]). *Assume that* f *is as in Theorem 1.1. If* $p^{(j)}, q^{(j)}$ $(j = 1, 2)$ *are probability distributions satisfying the condition,*

$$r \le \frac{p_i^{(j)}}{q_i^{(j)}} \le R \quad \textit{for all } \ i \in \{1, \dots, n\} \ \textit{ and } \ j \in \{1, 2\},$$

then, obviously,

$$r \le \frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} \le R \quad \textit{for all } i \in \{1, \dots, n\} \ \textit{ and } \ \lambda \in [0, 1]$$

and we have the inequality,

$$\begin{aligned} 0 &\le I_f\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &\quad - \lambda I_f\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) I_f\left(p^{(2)}, q^{(2)}\right) \\ &\le H \lambda^r (1-\lambda)^r \sum_{i=1}^n \frac{\left|\det \begin{bmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{bmatrix}\right|^r}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r} \\ &\quad \times \left[\lambda^{1-r} \left[q_i^{(1)}\right]^{1-r} + (1-\lambda)^{1-r} \left[q_i^{(2)}\right]^{1-r}\right] \end{aligned} \tag{1.7}$$

for all $\lambda \in [0, 1]$.

PROOF. We follow the proof in [**40**].

If we choose, in (1.2),

$$x = \frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} \quad \text{and} \quad y = \frac{p_i^{(1)}}{q_i^{(1)}} \ \ (i = 1, \dots, n)$$

we get,

$$\begin{aligned} &\left| f\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)\right| \\ &\le H \left|\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right|^r \\ &= H \frac{\left|\lambda q_i^{(1)} p_i^{(1)} + (1-\lambda) q_i^{(1)} p_i^{(2)} - \lambda p_i^{(1)} q_i^{(1)} - (1-\lambda) p_i^{(1)} q_i^{(2)}\right|^r}{\left[q_i^{(1)}\right]^r \left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r} \\ &= \frac{H (1-\lambda)^r \left|q_i^{(1)} p_i^{(2)} - p_i^{(1)} q_i^{(2)}\right|^r}{\left[q_i^{(1)}\right]^r \left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r}. \end{aligned} \tag{1.8}$$

If we multiply (1.8) by $\lambda q_i^{(1)}$, we obtain,

$$\left|\lambda q_i^{(1)} f\left(\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}\right)-\lambda q_i^{(1)} f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)\right| \leq \frac{H\lambda\,(1-\lambda)^r\left[q_i^{(1)}\right]^{1-r}\left|q_i^{(1)}p_i^{(2)}-p_i^{(1)}q_i^{(2)}\right|^r}{\left[\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}\right]^r} \tag{1.9}$$

for all $i\in\{1,\dots,n\}$ and $\lambda\in[0,1]$.

If in (1.2) we choose x as above but with

$$y=\frac{p_i^{(2)}}{q_i^{(2)}}\quad(i=1,\dots,n)\,,$$

we get,

$$\begin{aligned}\left|f\left(\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}\right)-f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right| &\leq H\left|\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right|^r\\ &= H\frac{\left|\lambda p_i^{(1)}q_i^{(2)}+(1-\lambda)\,p_i^{(2)}q_i^{(2)}-\lambda p_i^{(2)}q_i^{(1)}-(1-\lambda)\,q_i^{(2)}p_i^{(2)}\right|^r}{\left[\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}\right]^r\left[q_i^{(2)}\right]^r}\\ &= \frac{H\,(1-\lambda)^r\left|p_i^{(1)}q_i^{(2)}-p_i^{(2)}q_i^{(1)}\right|^r}{\left[q_i^{(2)}\right]^r\left[\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}\right]^r}.\end{aligned} \tag{1.10}$$

If we now multiply (1.10) by $(1-\lambda)\,q_i^{(2)}$, we obtain,

$$\left|(1-\lambda)\,q_i^{(2)} f\left(\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}\right)-(1-\lambda)\,q_i^{(2)} f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right| \leq \frac{H\lambda^r\,(1-\lambda)\left[q_i^{(2)}\right]^{1-r}\left|q_i^{(1)}p_i^{(2)}-p_i^{(1)}q_i^{(2)}\right|^r}{\left[\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}\right]^r} \tag{1.11}$$

for all $i\in\{1,\dots,n\}$ and $\lambda\in[0,1]$.

If we add (1.5) and (1.11) and use the triangle inequality, we get,

$$\begin{aligned}(1.12)\quad &\left|\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right] f\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right)\right. \\ &\qquad \left. -\lambda q_i^{(1)} f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) - (1-\lambda) q_i^{(2)} f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right| \\ &\leq \frac{H\lambda^r (1-\lambda)^r \left|\det\begin{bmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)}\end{bmatrix}\right|^r}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^r} \left[\lambda^{1-r}\left[q_i^{(1)}\right]^{1-r} + (1-\lambda)^{1-r}\left[q_i^{(2)}\right]^{1-r}\right]\end{aligned}$$

for all $i \in \{1, \dots, n\}$ and $\lambda \in [0,1]$.

Summing (1.12) over i from 1 to n and using the generalised triangle inequality, we obtain the desired inequality (1.7). ∎

REMARK 1.2. If we assume that f is $L-$Lipschitzian, then the inequality (1.7) becomes,

$$\begin{aligned}(1.13)\quad 0 &\leq I_f\left(\lambda p^{(1)} + (1-\lambda) p^{(2)}, \lambda q^{(1)} + (1-\lambda) q^{(2)}\right) \\ &\quad - \lambda I_f\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) I_f\left(p^{(2)}, q^{(2)}\right) \\ &\leq L\lambda(1-\lambda)\sum_{i=1}^n \frac{\left|\det\begin{bmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)}\end{bmatrix}\right|}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}},\end{aligned}$$

for all $\lambda \in [0,1]$.

2. SOME APPLICATIONS FOR PARTICULAR DIVERGENCES

Using the inequality (1.6), i.e.,

$$(2.1)\quad 0 \leq I_f(p,q) \leq \|f'\|_\infty V(p,q),$$

provided that f is absolutely continuous on $[r,R]$ and $f' \in L_\infty[r,R]$, $\|f'\|_\infty := ess \sup_{t\in[r,R]} |f'(t)|$, we are able to point out a number of additional inequalities between different divergence measures.

PROPOSITION 2.1. *Let p, q be two probability distributions with the property that,*

$$(2.2)\quad 0 < r \leq \frac{p_i}{q_i} =: r_i < R < \infty \text{ for all } i \in \{1, \dots, n\},$$

then,

$$(2.3)\quad \begin{aligned} 0 &\leq KL(p,q) \\ &\leq \begin{cases} \left[\ln\sqrt{\frac{R}{r}} + \left|1 + \ln\sqrt{rR}\right|\right] V(p,q) & \text{if } 0 < r \leq e^{-1} \\ (1 + \ln R) V(p,q) & \text{if } e^{-1} < r < 1 \end{cases}.\end{aligned}$$

PROOF. Consider the mapping $f : (0,\infty) \to \mathbb{R}$, $f(t) = t\ln t$, then

$$f'(t) = \ln(et).$$

It follows that,

$$\|f'\|_\infty = \sup_{t\in[r,R]} |f'(t)| = \max\{|\ln(er)|, \ln(eR)\}.$$

We note the following.

(1) If $0 < r \le e^{-1}$, then $|\ln(er)| = \ln(er) = -1 - \ln r$ and

$$\max\{|\ln(er)|, \ln(eR)\} = \frac{-1 - \ln r + \ln R + 1 + |\ln r + 1 + 1 + \ln R|}{2}$$
$$= \ln\left(\frac{R}{2}\right)^{\frac{1}{2}} + \left|1 + \ln\sqrt{rR}\right|.$$

(2) If $e^{-1} < r < 1$, then $|\ln(er)| = 1 + \ln r$ and

$$\max\{|\ln(er)|, \ln(eR)\} = \max\{\ln(er), \ln(eR)\} = 1 + \ln R$$

and the proposition is proved.

∎

PROPOSITION 2.2. *Let p, q be as in Proposition 2.1, then we have the inequality:*

$$0 \le KL(q,p) \le \frac{1}{r} V(p,q). \tag{2.4}$$

PROOF. Consider the mapping $f(t) = -\ln t$, $t \in (0,\infty)$, then,

$$f'(t) = -\frac{1}{t} \quad \text{and} \quad \|f'\|_\infty = \frac{1}{r}.$$

Since,

$$I_f(p,q) = -\sum_{i=1}^{n} q_i \ln\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{n} q_i \ln\left(\frac{q_i}{p_i}\right) = KL(q,p),$$

(2.1) gives the desired result (2.4). ∎

We point out now a bound for χ^2-divergence.

PROPOSITION 2.3. *Let p, q be as above, then,*

$$0 \le D_{\chi^2}(p,q) \le 2(R-1) V(p,q). \tag{2.5}$$

PROOF. Consider the mapping $f : (0,\infty) \to \mathbb{R}$, $f(t) = (t-1)^2$. As $f'(t) = 2(t-1)$, it follows that $\|f'\|_\infty = 2(R-1)$. Using (2.1), we obtain (2.5). ∎

The following result for Hellinger discrimination also holds.

PROPOSITION 2.4. *Assuming that the probability distributions p, q satisfy the condition (2.1), then,*

$$0 \le h^2(p,q) \le \frac{1}{4}\left[\frac{\sqrt{R}-\sqrt{r}}{\sqrt{rR}} + \left|\frac{\sqrt{R}+\sqrt{r}}{\sqrt{rR}} - 2\right|\right] V(p,q). \tag{2.6}$$

PROOF. Consider the mapping $f : (0,\infty) \to \mathbb{R}$, $f(t) = \frac{1}{2}\left(\sqrt{t}-1\right)^2$, then,

$$f'(t) = \frac{1}{2} - \frac{1}{2\sqrt{t}} \quad \text{and} \quad f''(t) = \frac{1}{4\sqrt{t^3}},\ t \in (0,\infty).$$

It follows that,

$$\begin{aligned}\|f'\|_\infty &= \sup_{t\in[r,R]}|f'(t)| = \max\left\{\left|\frac{\sqrt{r}-1}{2\sqrt{r}}\right|,\left|\frac{\sqrt{R}-1}{2\sqrt{R}}\right|\right\}\\ &= \max\left\{\frac{1-\sqrt{r}}{2\sqrt{r}},\frac{\sqrt{R}-1}{2\sqrt{R}}\right\}\\ &= \frac{1}{4}\left[\frac{\sqrt{R}-\sqrt{r}}{\sqrt{rR}}+\left|2-\frac{\sqrt{R}+\sqrt{r}}{\sqrt{rR}}\right|\right].\end{aligned}$$

Using (2.1), we deduce (2.6). ∎

Now, consider Bhattacharyya distance.

PROPOSITION 2.5. *Assuming that p, q are probability distributions, then,*

$$0 \le 1 - B(p,q) \le V_{\frac{1}{2}}(p,q), \tag{2.7}$$

where $V_{\frac{1}{2}}(p,q) = \sum_{i=1}^{n}\sqrt{q_i(p_i-q_i)}$ is the $\frac{1}{2}-$variational distance.

PROOF. Consider the mapping $f:[0,\infty)\to[0,\infty)$ given by,

$$f(t) = -\sqrt{t}+1.$$

Obviously,

$$|f(x)-f(y)| = \left|\sqrt{x}-\sqrt{y}\right| \le \sqrt{|x-y|}, \quad \text{for all } x,y\in[0,\infty),$$

which shows that f is of the $\frac{1}{2}-$Hölder type with constant $H=1$.
Applying Theorem 1.1, we deduce (2.7). ∎

Another inequality for Bhattacharyya distance in terms of the variational distance V is embodied in the following proposition.

PROPOSITION 2.6. *Assuming that the probability distributions p, q satisfy the the condition:*

$$0 < r \le \frac{p_i}{q_i} \quad \textit{for all } \ i\in\{1,\dots,n\}, \tag{2.8}$$

then,

$$0 \le 1 - B(p,q) \le \frac{1}{2\sqrt{r}}V(p,q). \tag{2.9}$$

PROOF. For the mapping $f(t) = -\sqrt{t}+1$, we have $f'(t) = -\frac{1}{2\sqrt{t}}$ and $\|f'\|_\infty = \sup_{t\in[0,\infty)} = \frac{1}{2\sqrt{r}}$. Applying (2.1), we deduce (2.9). ∎

PROPOSITION 2.7. *Assuming that the probability distributions p, q satisfy the condition (2.8), then,*

$$0 \le 1 - M(p,q) \le \frac{2}{(r+1)^2}V(p,q), \tag{2.10}$$

where $M(p,q)$ is harmonic distance.

PROOF. Consider the mapping $f : [0,\infty) \to \mathbb{R}$, $f(t) = 1 - \frac{2t}{t+1}$. Obviously, $f'(t) = -\frac{2}{(t+1)^2}$, $f''(t) = \frac{4}{(t+1)^3}$ and $\|f'\|_\infty = \sup_{t\in[r,\infty)} |f'(t)| = \frac{2}{(r+1)^2}$. As

$$I_f(p,q) = 1 - M(p,q),$$

then by (2.1) we deduce (2.10). ∎

PROPOSITION 2.8. *Assuming that the probability distributions p, q satisfy the condition:*

$$\frac{p_i}{q_i} \le R < \infty, \quad (i = 1,\dots,n), \tag{2.11}$$

then,

$$0 \le J(p,q) \le \left(\ln R - \frac{1}{R} + 1\right) V(p,q), \tag{2.12}$$

where J is Jeffreys' divergence.

PROOF. Consider the mapping $f(t) = (t-1)\ln t$, $t > 0$, then,

$$f'(t) = \ln t - \frac{1}{t} + 1; \quad t \in (0,\infty),$$
$$f''(t) = \frac{t+1}{t^2}, \quad t \in (0,\infty),$$

hence,

$$\|f'\|_\infty = \sup_{t\in(0,R]} |f'(t)| = f'(R) = \ln R - \frac{1}{R} + 1.$$

As

$$I_f(p,q) = J(p,q),$$

then by (2.1) we deduce (2.12). ∎

Finally, the following result for triangular discrimination holds.

PROPOSITION 2.9. *If p, q are such that the condition (2.11) holds, then,*

$$0 \le \Delta(p,q) \le \frac{(R-1)(R+3)}{(R+1)^2} V(p,q) \le V(p,q). \tag{2.13}$$

The proof is obvious by (2.2) applied for the mapping $f(t) = \frac{(t-1)^2}{t+1}$.

CHAPTER 6

Inequalities for Two f-Divergences

In this chapter two genereal f-divergence measures are compared. Cauchy mean-value theorem is empolyed and applications for various particular inequalities are provided.

1. Some General Estimates

We start with the following result [**38**].

THEOREM 1.1 (Dragomir, 2001 [**38**]). *Let $f, g : [0, \infty) \to \mathbb{R}$ be two mappings such that $f(1) = g(1) = 0$. If there exist real constants m, M such that,*

$$m|f(x) - f(y)| \leq |g(x) - g(y)| \leq M|f(x) - f(y)| \tag{1.1}$$

for all $x, y \in [r, R] \subset (0, \infty)$, then,

$$mI_{|f|}(p, q) \leq I_{|g|}(p, q) \leq MI_{|f|}(p, q) \tag{1.2}$$

for all p, q probability distributions with $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ for all $i \in \{1, \dots, n\}$.

PROOF. By (1.1) it follows that,

$$\begin{aligned} m\left|f\left(\frac{p_i}{q_i}\right)\right| &= m\left|f\left(\frac{p_i}{q_i}\right) - f(1)\right| \leq \left|g\left(\frac{p_i}{q_i}\right) - g(1)\right| \\ &= \left|g\left(\frac{p_i}{q_i}\right)\right| \leq M\left|f\left(\frac{p_i}{q_i}\right) - f(1)\right| = M\left|f\left(\frac{p_i}{q_i}\right)\right| \end{aligned} \tag{1.3}$$

for all $i \in \{1, \dots, n\}$.

If we multiply (1.3) by $q_i \geq 0$ and sum the obtained inequalities, we deduce (1.2). ∎

COROLLARY 1.2 (Dragomir, 2001 [**38**]). *Assume that the mappings $f, g : [0, \infty) \to \mathbb{R}$ are as above and that f, g are differentiable on (r, R) with $f'(t) \neq 0$ for $t \in (r, R)$ and*

$$-\infty < m = \inf_{t \in (r,R)} \left|\frac{g'(t)}{f'(t)}\right|, \qquad \sup_{t \in (r,R)} \left|\frac{g'(t)}{f'(t)}\right| = M < \infty, \tag{1.4}$$

then we have the inequality (1.2) for all p, q as above.

PROOF. We use the following Cauchy theorem:

If $\gamma, f : [a, b] \to \mathbb{R}$ are continuous and differentiable on (a, b) and $f'(t) \neq 0$ for all $t \in (a, b)$, then there exists $c \in [a, b]$ such that,

$$\frac{\gamma(b) - \gamma(a)}{f(b) - f(a)} = \frac{\gamma'(c)}{f'(c)}.$$

Now, suppose that $x, y \in [r, R]$ and $x < y$, then, by Cauchy's theorem, we have,

$$m \leq \left|\frac{g(x) - g(y)}{f(x) - f(y)}\right| = \left|\frac{g'(z)}{f'(z)}\right| \leq M$$

and we can conclude that for any $x, y \in [r, R]$ we have,

$$m|f(x) - f(y)| \leq |g(x) - g(y)| \leq M|f(x) - f(y)|.$$

Applying Theorem 1.1, we deduce (1.2). ∎

The following corollary for the variational distance holds.

COROLLARY 1.3. *Let* $g : [0, \infty) \to \mathbb{R}$ *be a mapping such that* $g(1) = 0$*. If there exist real constants* n, N *such that,*

$$n|x-y| \le |g(x) - g(y)| \le N|x-y| \quad \text{for all } x, y \in [r, R], \tag{1.5}$$

then,

$$nV(p,q) \le I_{|g|}(p,q) \le NV(p,q) \tag{1.6}$$

for any probability distributions p, q *with* $0 < r \le \frac{p_i}{q_i} \le R < \infty$ *for all* $i \in \{1, \dots, n\}$*.*

The proof is obvious by Theorem 1.1, choosing $f(x) = x - 1$.

COROLLARY 1.4. *Assuming that the mapping* g *is continuous on* $[a, b]$ *and differentiable on* (a, b) *and*

$$-\infty < n = \inf_{t \in (r,R)} |g'(t)|, \qquad \sup_{t \in (r,R)} |g'(t)| = N < \infty,$$

then we have the inequality (1.6) for all p, q *as above.*

2. PARTICULAR CASES IN TERMS OF THE VARIATIONAL DISTANCE

PROPOSITION 2.1. *Let* $0 < r \le \frac{p_i}{q_i} \le R < \infty$ $(i = 1, \dots, n)$*, then we have the inequality,*

$$0 \le KL(p,q) \le \begin{cases} [\ln R + 1] V(p,q) & \text{if } r \ge e^{-1}, \\ \\ \max\{\ln R + 1; |\ln R + 1|\} V(p,q) & \text{if } r < e^{-1}. \end{cases} \tag{2.1}$$

PROOF. Consider the mapping $g : (0, \infty) \to \mathbb{R}$, $g(t) = t \ln t$, then $g'(t) = \ln t + 1$ and obviously,

$$M := \sup_{t \in (r,R)} |g'(t)| = \begin{cases} \ln R + 1 & \text{if } r \ge e^{-1}, \\ \\ \max\{\ln R + 1; |\ln R + 1|\} & \text{if } r < e^{-1}. \end{cases}$$

Applying Corollary 1.4, we can state,

$$\sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} \ln\left(\frac{p_i}{q_i}\right) \right| \le NV(p,q).$$

By the generalised triangle inequality, we have,

$$\begin{aligned} KL(p,q) &= \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right) = \left| \sum_{i=1}^{n} p_i \ln\left(\frac{p_i}{q_i}\right) \right| \\ &\le \sum_{i=1}^{n} p_i \left| \ln\left(\frac{p_i}{q_i}\right) \right| \le NV(p,q) \end{aligned}$$

and the inequality (2.1) is proved. ∎

If we introduce the *modified Kullback-Leibler* distance,

$$|KL|(p,q) = \sum_{i=1}^{n} p_i \left| \ln\left(\frac{p_i}{q_i}\right) \right|,$$

then obviously,

$$K(p,q) \le |KL|(p,q) \quad \text{for all } p, q \in \mathbb{P}^n. \tag{2.2}$$

For this modified distance, we may prove the following as well.

PROPOSITION 2.2. *Let* $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ $(i = 1, \dots, n)$*, then,*

$$(\ln r + 1) V(p, q) \leq |KL|(p, q) \leq (\ln R + 1) V(p, q), \tag{2.3}$$

provided that $r \geq e^{-1}$*.*

PROOF. The second inequality in (1.3) has been proven above.

For the first inequality, we can apply Corollary 1.4 by observing that for $g(t) = t \ln t$, and $r \geq e^{-1}$,

$$\inf_{t \in [r,R]} |g'(t)| = \ln r + 1.$$

We omit the details. ∎

PROPOSITION 2.3. *Let* $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ $(i = 1, \dots, n)$*, then,*

$$KL(q, p) \leq \frac{1}{r} V(p, q). \tag{2.4}$$

PROOF. Consider the mapping $g : (0, \infty) \to \mathbb{R}$, $g(t) = \ln t$, then $g'(t) = \frac{1}{t}$ and obviously,

$$M := \sup_{t \in [r,R]} |g'(t)| = \frac{1}{r}.$$

Applying Corollary 1.3, we can state:

$$\sum_{i=1}^{n} q_i \left| \ln \left(\frac{p_i}{q_i} \right) \right| \leq \frac{1}{r} V(p, q).$$

By the generalised triangle inequality, we have,

$$\begin{aligned} K(q, p) = \sum_{i=1}^{n} q_i \ln \left(\frac{q_i}{p_i} \right) &= \left| \sum_{i=1}^{n} q_i \ln \left(\frac{q_i}{p_i} \right) \right| \\ &\leq \sum_{i=1}^{n} q_i \left| \ln \left(\frac{p_i}{q_i} \right) \right| \leq \frac{1}{r} V(p, q) \end{aligned}$$

and the proposition is proved. ∎

The following result for the modified Kullback-Leibler distance also holds.

PROPOSITION 2.4. *Let* p, q *be as in Proposition 2.3, then,*

$$\frac{1}{R} V(p, q) \leq |KL|(q, p) \leq \frac{1}{r} V(p, q). \tag{2.5}$$

PROOF. The second inequality in (2.5) has been proven above. The first inequality follows by the first inequality in Corollary 1.4 by taking into account that,

$$m = \inf_{t \in (r,R)} |g'(t)| = \frac{1}{R}.$$

∎

The following result for *Hellinger discrimination* holds.

PROPOSITION 2.5. *Let* $0 < r \leq \frac{p_i}{q_i} \leq R < \infty \ (i = 1, \dots, n)$, *then,*

$$\left[\frac{\sqrt{R}-\sqrt{r}}{4\sqrt{rR}} - \left|\frac{\sqrt{R}+\sqrt{r}}{4\sqrt{rR}} - \frac{1}{2}\right|\right] V(p,q) \leq h^2(p,q) \leq \left[\frac{\sqrt{R}-\sqrt{r}}{4\sqrt{rR}} + \left|\frac{\sqrt{R}+\sqrt{r}}{4\sqrt{rR}} - \frac{1}{2}\right|\right] V(p,q). \tag{2.6}$$

PROOF. Consider the mapping $g : (0, \infty) \to \mathbb{R}$, $g(t) = \frac{1}{2}\left(\sqrt{t} - 1\right)^2$, then,

$$g'(t) = \frac{1}{2} \cdot \frac{\sqrt{t}-1}{\sqrt{t}}, \quad t \in (0, \infty),$$

$$\begin{aligned} n &= \inf_{t \in [r,R]} |g'(t)| = \min\{|g'(r)|, |g'(R)|\} \\ &= \frac{|g'(r)| + |g'(R)| - ||g'(r)| - |g'(R)||}{2} \\ &= \frac{\sqrt{R}-\sqrt{r}}{4\sqrt{rR}} - \left|\frac{\sqrt{R}+\sqrt{r}}{4\sqrt{rR}} - \frac{1}{2}\right| \end{aligned}$$

and

$$\begin{aligned} N &= \sup_{t \in [r,R]} |g'(t)| = \max\{|g'(r)|, |g'(R)|\} \\ &= \frac{|g'(r)| + |g'(R)| + ||g'(r)| - |g'(R)||}{2} \\ &= \frac{\sqrt{R}-\sqrt{r}}{4\sqrt{rR}} + \left|\frac{\sqrt{R}+\sqrt{r}}{4\sqrt{rR}} - \frac{1}{2}\right| \end{aligned}$$

respectively.

As $g(t) \geq 0$, then obviously,

$$I_{|g|}(p,q) = I_g(p,q) = h^2(p,q).$$

Using (1.6), we obtain (2.6). ∎

REMARK 2.1. The inequality (2.6) is equivalent to,

$$\left|h^2(p,q) - \frac{\sqrt{R}-\sqrt{r}}{4\sqrt{rR}} V(p,q)\right| \leq \left|\frac{\sqrt{R}+\sqrt{r}}{4\sqrt{rR}} - \frac{1}{2}\right| V(p,q). \tag{2.7}$$

We now point out some inequalities for *chi-square distance.*

PROPOSITION 2.6. *Let* $0 < r \leq \frac{p_i}{q_i} \leq R < \infty \ (i = 1, \dots, n)$*, then,*

$$\begin{aligned} [R - r - |R + r - 2|] V(p,q) &\leq D_{\chi^2}(p,q) \\ &\leq [R - r + |R + r - 2|] V(p,q). \end{aligned} \tag{2.8}$$

PROOF. If $g : (0, \infty) \to \mathbb{R}$, $g(t) = (t-1)^2$, then, $g'(t) = 2(t-1)$,

$$\begin{aligned} n &= \inf_{t \in [r,R]} |g'(t)| = \min\{|g'(r)|, |g'(R)|\} \\ &= R - r - |R + r - 2| \end{aligned}$$

and

$$N = R - r + |R + r - 2|.$$

Using (1.6), and taking into account that $g(t) \geq 0$, $t \in \mathbb{R}$, and

$$I_g(p,q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} = D_{\chi^2}(p,q),$$

we deduce (2.8). ∎

REMARK 2.2. The inequality (2.8) is equivalent to,

$$|D_{\chi^2}(p,q) - (R-r)V(p,q)| \leq |R+r-2|\, V(p,q). \tag{2.9}$$

We point out now some inequalities for the *Bhattacharyya distance.*

PROPOSITION 2.7. *Let* $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$ $(i = 1, \dots, n)$*, then,*

$$0 \leq 1 - B(p,q) \leq \frac{1}{2\sqrt{r}} V(p,q). \tag{2.10}$$

PROOF. Consider the mapping $g(t) = 1 - \sqrt{t}$, $t \in (0, \infty)$, then $g(1) = 0$, $g'(t) = -\frac{1}{2\sqrt{t}}$ and

$$N = \sup_{t \in [r,R]} |g'(t)| = \sup_{t \in [r,R]} \frac{1}{2\sqrt{t}} = \frac{1}{2\sqrt{r}}.$$

Applying Corollary 1.4,

$$\sum_{i=1}^{n} q_i \left|1 - \sqrt{\frac{p_i}{q_i}}\right| \leq \frac{1}{2\sqrt{r}} V(p,q),$$

which is equivalent to

$$\sum_{i=1}^{n} |q_i - \sqrt{p_i q_i}| \leq \frac{1}{2\sqrt{r}} V(p,q). \tag{2.11}$$

Using the generalised triangle inequality, we obtain,

$$\begin{aligned} \sum_{i=1}^{n} |q_i - \sqrt{p_i q_i}| &\geq \left| \sum_{i=1}^{n} (q_i - \sqrt{p_i q_i}) \right| \\ &= |1 - B(p,q)| = 1 - B(p,q). \end{aligned}$$

∎

If we define the following distance $\tilde{B}(p,q) := \sum_{i=1}^{n} \sqrt{q_i} \left|\sqrt{q_i} - \sqrt{p_i}\right|$, then we may state the following proposition as well.

PROPOSITION 2.8. *Assume that* p_i, q_i, r, R *are as above, then,*

$$\frac{1}{2\sqrt{R}} V(p,q) \leq \tilde{B}(p,q) \leq \frac{1}{2\sqrt{r}} V(p,q). \tag{2.12}$$

The proof is obvious by Corollary 1.3 applied for the mapping $g(t) = 1 - \sqrt{t}$.

PROPOSITION 2.9. *Assuming that* p_i, q_i, r, R *are as above, then,*

$$0 \leq 1 - M(p,q) \leq \frac{2}{(r+1)^2} V(p,q), \tag{2.13}$$

where $M(p,q)$ *is harmonic distance.*

PROOF. Consider the mapping $g(t) = 1 - \frac{2t}{t+1}$, then, $g(1) = 0$, $g'(t) = -\frac{2}{(t+1)^2}$ and

$$N := \sup_{t \in [r,R]} |g'(t)| = \frac{2}{(r+1)^2}.$$

Applying Corollary 1.4, we can state that,

$$\sum_{i=1}^{n} q_i \left| 1 - \frac{2\frac{p_i}{q_i}}{\frac{p_i}{q_i} + 1} \right| \leq \frac{2}{(r+1)^2} V(p,q),$$

which is clearly equivalent to:

$$\sum_{i=1}^{n} \frac{q_i |p_i - q_i|}{p_i + q_i} \leq \frac{2}{(r+1)^2} V(p,q). \tag{2.14}$$

Using the generalised triangle inequality, we obtain (2.13). ∎

If we introduce the divergence measure:

$$\tilde{M}(p,q) := \sum_{i=1}^{n} q_i \cdot \frac{|p_i - q_i|}{p_i + q_i} = I_l(p,q),$$

where $l(t) = \frac{|t-1|}{t+1}$, $t > 0$, then we have the following proposition.

PROPOSITION 2.10. *With the above assumptions,*

$$\frac{2}{(R+1)^2} V(p,q) \leq \tilde{M}(p,q) \leq \frac{2}{(r+1)^2} V(p,q). \tag{2.15}$$

Finally, consider *Jeffreys' distance*.

PROPOSITION 2.11. *Assume that* $0 < r \leq \frac{p_i}{q_i} \leq R < \infty$*, then,*

$$\begin{aligned} \left[\frac{R-r}{2rR} + \ln\sqrt{\frac{R}{r}} - \left| \frac{R-r}{2rR} - \ln\sqrt{rR} - 1 \right| \right] V(p,q) \\ \leq J(p,q) \leq \left[\frac{R-r}{2rR} + \ln\sqrt{\frac{R}{r}} + \left| \frac{R-r}{2rR} - \ln\sqrt{rR} - 1 \right| \right] V(p,q). \end{aligned} \tag{2.16}$$

PROOF. Consider the mapping $g(t) = (t-1)\ln t$, $t > 0$, then, $g'(t) = \ln t - \frac{1}{t} + 1$, $g''(t) = \frac{t+1}{t^2}$, which shows that $g'(\cdot)$ is strictly increasing on $(0,\infty)$ and $g'(1) = 0$. It can be seen that,

$$\begin{aligned} n &= \inf_{t \in [r,R]} |g'(t)| = \min\{|g'(r)|, |g'(R)|\} \\ &= \frac{R-r}{2rR} + \ln\sqrt{\frac{R}{r}} - \left| \frac{R+r}{2rR} - \ln\sqrt{rR} - 1 \right| \end{aligned}$$

and

$$\begin{aligned} N &= \sup_{t \in [r,R]} |g'(t)| = \max\{|g'(r)|, |g'(R)|\} \\ &= \frac{R-r}{2rR} + \ln\sqrt{\frac{R}{r}} + \left| \frac{R+r}{2rR} - \ln\sqrt{rR} - 1 \right|. \end{aligned}$$

In addition, as,

$$I_{|g|}(p,q) = \sum_{i=1}^{n} q_i \left| \left(\frac{p_i}{q_i} - 1 \right) \right| \left| \ln \left(\frac{p_i}{q_i} \right) \right| = \sum_{i=1}^{n} |p_i - q_i| \, |\ln p_i - \ln q_i|$$
$$= \sum_{i=1}^{n} (p_i - q_i)(\ln p_i - \ln q_i) = I(p,q),$$

then by (1.6), we deduce (2.16). ∎

REMARK 2.3. The above inequality (2.16) is equivalent to

$$\left| J(p,q) - \left[\frac{R-r}{2rR} + \ln \sqrt{\frac{R}{r}} \right] V(p,q) \right| \leq \left| \frac{R+r}{2rR} - \ln \sqrt{rR} - 1 \right| V(p,q). \tag{2.17}$$

3. OTHER PARTICULAR CASES

Consider the modified Kullback-Leibler divergence.

PROPOSITION 3.1. *Assume that* $0 < r \leq \frac{p_i}{q_i} \leq R < \infty \ (i = 1, \ldots, n)$*, then,*

$$0 \leq KL(p,q) \leq \left[\frac{R-r}{2} + \ln \sqrt{\frac{R^R}{r^r}} + \left| \frac{r+R}{2} + \ln \sqrt{R^R r^r} \right| \right] |KL|(q,p). \tag{3.1}$$

PROOF. Consider the mappings $g(t) = t \ln t$, $f(t) = \ln t$, $t > 0$ and $h(t) := \frac{g'(t)}{f'(t)} = t \ln t + t$.

We observe that,

$$M = \sup_{t \in [r,R]} |h(t)| = \max \{ |h(r)|, |h(R)| \}$$
$$= \frac{R-r}{2} + \ln \sqrt{\frac{R^R}{r^r}} + \left| \frac{r+R}{2} + \ln \sqrt{R^R r^r} \right|.$$

Applying Corollary 1.2,

$$\sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} \ln \left(\frac{p_i}{q_i} \right) \right| \leq M \sum_{i=1}^{n} q_i \left| \ln \left(\frac{p_i}{q_i} \right) \right| = M \, |KL|(q,p)$$

and since, by the generalised triangle inequality, we have,

$$\sum_{i=1}^{n} p_i \left| \ln \left(\frac{p_i}{q_i} \right) \right| \geq |KL(p,q)| = KL(p,q) \geq 0,$$

and the inequality (3.1) is proved. ∎

We now compare the Hellinger discrimination with $|KL|$.

PROPOSITION 3.2. *Let p_i, q_i, r, R be as in Proposition 3.1, then,*

$$\frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}-\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right|\right]|KL|(q,p) \tag{3.2}$$
$$\leq h^2(p,q) \leq \frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}+\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right|\right]|KL|(q,p).$$

PROOF. Consider the mappings $g(t)=\frac{1}{2}\left(\sqrt{t}-1\right)^2$, $f(t)=\ln t$, $t>0$ with

$$h(t):=\frac{g'(t)}{f'(t)}=\frac{1}{2}\left(\frac{\sqrt{t}-1}{\sqrt{t}}\right)\cdot t=\frac{1}{2}\left(\sqrt{t}-1\right)\sqrt{t},\ t>0.$$

We observe that,

$$\begin{aligned} m &= \inf_{t\in[r,R]}|h(t)| = \min\{|h(r)|,|h(R)|\} \\ &= \frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}-\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right|\right]\end{aligned}$$

and, analogously,

$$\begin{aligned} M &= \sup_{t\in[r,R]}|h(t)| = \max\{|h(r)|,|h(R)|\} \\ &= \frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}+\left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right|\right].\end{aligned}$$

Since $g(t)\geq 0$, we have,

$$I_{|g|}(p,q)=I_g(p,q)=h^2(p,q)$$

and then, by Corollary 1.2, we deduce (3.2). ∎

REMARK 3.1. The above inequality is equivalent to

$$\left|h^2(p,q)-\frac{1}{2}\left[\frac{R-r}{2}-\frac{\sqrt{R}-\sqrt{r}}{2}\right]|KL(q,p)|\right| \tag{3.3}$$
$$\leq \left|\frac{\sqrt{r}+\sqrt{R}}{2}-\frac{r+R}{2}\right||KL|(q,p).$$

We now compare the Chi-square distance with $|KL|$.

PROPOSITION 3.3. *Let p_i, q_i, r, R be as above, then,*

$$\left[(R-r)(R+r-1)-\left|R+r-\left(R^2+r^2\right)\right|\right]|KL|(q,p) \tag{3.4}$$
$$\leq D_{\chi^2}(p,q) \leq \left[(R-r)(R+r-1)+\left|R+r-\left(R^2+r^2\right)\right|\right]|KL|(q,p).$$

PROOF. Consider the mappings $g(t) = (t-1)^2$, $f(t) = \ln t$, $t > 0$, with $h(t) = \frac{g'(t)}{f'(t)} = 2t(t-1)$.

$$
\begin{aligned}
m = \inf_{t\in[r,R]} |h(t)| &= \frac{1}{2}\left[2r(1-r) + 2R(R-1) - |2r(1-r) - 2R(R-1)|\right] \\
&= \left[r - r^2 + R^2 - R - \left|r - r^2 - R^2 + R\right|\right] \\
&= R^2 - r^2 - (R-r) - \left|R + r - \left(R^2 + r^2\right)\right| \\
&= (R-r)(R+r-1) - \left|R + r - \left(R^2 + r^2\right)\right|
\end{aligned}
$$

and

$$M = \sup_{t\in[r,R]} (h(t)) = (R-r)(R+r-1) + \left|R + r - \left(R^2 + r^2\right)\right|.$$

Since $g(t) \geq 0$, we have,

$$I_{|g|}(p,q) = I_g(p,q) = D_{\chi^2}(p,q)$$

and by Corollary 1.2, we deduce (3.4). ∎

REMARK 3.2. The above inequality is equivalent to

$$
\begin{aligned}
(3.5)\qquad &|D_{\chi^2}(p,q) - (R-r)(R+r-1)|KL|(q,p)| \\
&\leq \left|R + r - \left(R^2 + r^2\right)\right| |KL|(q,p).
\end{aligned}
$$

CHAPTER 7

Some Inequalities for Lipschitzian Mappings

In this chapter various Jensen's type inequalities for Lipschitzian functions with values in normed spaces are given. They are applied to norm inequalities and to f-divergence measure. A plethora of particular examples are provided involving various divergence measures considered above.

1. Some Analytic Inequalities

The following general result for Lipschitzian functions can be stated [**47**]:

THEOREM 1.1 (Dragomir, 2004 [**47**]). *Let X,Y be two normed linear spaces with the norms $\|\cdot\|$ and $|\cdot|$ respectively. If $F : X \to Y$ is L - Lipschitzian, that is,*

$$|F(x) - F(y)| \leq L\|x-y\| \quad \textit{for all } x, y \in X, \tag{1.1}$$

then for all $x_i \in X$, $p_i \geq 0$ with $\sum_{i=1}^n p_i = 1$ $(i = 1, \ldots, n)$, we have,

$$\left|F\left(\sum_{i=1}^n p_i x_i\right) - \sum_{i=1}^n p_i F(x_i)\right| \leq L \max_{k=1,\ldots,n-1} \|\Delta x_k\| \sum_{i,j=1}^n p_i p_j |i-j|. \tag{1.2}$$

PROOF. We follow the proof in [**47**].

As F is $L-$Lipschitzian, we can choose $x = \sum_{i=1}^n p_i x_i$ and $y = x_j$ $(j = 1, \ldots, n)$ in (1.1) to get,

$$\begin{aligned}\left|F\left(\sum_{i=1}^n p_i x_i\right) - F(x_j)\right| &\leq L\left\|\sum_{i=1}^n p_i x_i - x_j\right\| \\ &= L\left\|\sum_{i=1}^n p_i(x_i - x_j)\right\| \leq L\sum_{i=1}^n p_i\|x_i - x_j\|.\end{aligned} \tag{1.3}$$

Multiplying (1.3) by $p_j \geq 0$ and summing over j from 1 to n, we deduce,

$$\sum_{j=1}^n p_j \left|F\left(\sum_{i=1}^n p_i x_i\right) - F(x_j)\right| \leq L\sum_{i,j=1}^n p_i p_j \|x_j - x_i\|. \tag{1.4}$$

By the generalised triangle inequality we have,

$$\sum_{j=1}^n p_j \left|F\left(\sum_{i=1}^n p_i x_i\right) - F(x_j)\right| \geq \left|F\left(\sum_{i=1}^n p_i x_i\right) - \sum_{j=1}^n p_j F(x_j)\right|. \tag{1.5}$$

It is apparent that,

$$\sum_{i,j=1}^n p_i p_j \|x_i - x_j\| = 2\sum_{1\leq i<j\leq n} p_i p_j \|x_j - x_i\| \tag{1.6}$$

and that, for $i < j$,

$$x_j - x_i = \sum_{k=i}^{j-1} \Delta x_k,$$

where $\Delta x_k := x_{k+1} - x_k$ is the forward difference.

Applying the generalised triangle inequality we have,

$$\begin{aligned}(1.7)\qquad & \sum_{1\le i<j\le n} p_i p_j \|x_j - x_i\| \\ &= \sum_{1\le i<j\le n} p_i p_j \left\| \sum_{k=i}^{j-1} \Delta x_k \right\| \le \sum_{1\le i<j\le n} p_i p_j \sum_{k=i}^{j-1} \|\Delta x_k\| \\ &\le \sum_{1\le i<j\le n} p_i p_j (j-i) \max_{k=1,\dots,n-1} \|\Delta x_k\| \\ &= \max_{k=1,\dots,n-1} \|\Delta x_k\| \sum_{1\le i<j\le n} p_i p_j (j-i) \\ &= \frac{1}{2} \max_{k=1,\dots,n-1} \|\Delta x_k\| \sum_{i,j=1}^{n} p_i p_j |j-i| .\end{aligned}$$

Using (1.4) - (1.7) we deduce (1.2). ∎

COROLLARY 1.2 (Dragomir, 2004 [**47**]). *With the above assumptions for F and x_i $(i = 1, \dots, n)$, we have,*

$$(1.8)\qquad \left| F\left(\frac{1}{n} \sum_{i=1}^{n} x_i \right) - \frac{1}{n} \sum_{i=1}^{n} F(x_i) \right| \le L \cdot \frac{n^2-1}{3n} \max_{k=1,\dots,n-1} \|\Delta x_k\| .$$

PROOF. We choose $p_i = \frac{1}{n}$ $(i = 1, \dots, n)$ in (1.2) and compute

$$I := \sum_{i,j=1}^{n} |i-j| .$$

Observing that,

$$\begin{aligned}\sum_{j=1}^{n} |i-j| &= \sum_{j=1}^{i} |i-j| + \sum_{j=i+1}^{n} |i-j| = \sum_{j=1}^{i} (i-j) + \sum_{j=i+1}^{n} (j-i) \\ &= i^2 - \frac{i(i+1)}{2} + \sum_{j=1}^{n} j - i(n-i) \\ &= i^2 - (n+1)i + \frac{n(n+1)}{2} = \frac{n^2-1}{4} + \left(i - \frac{n+1}{2} \right)^2 .\end{aligned}$$

It follows that,

$$\begin{aligned}I &= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |i-j| \right) = \sum_{i=1}^{n} \left[i^2 - (n+1)i + \frac{n(n+1)}{2} \right] \\ &= \frac{(n-1)n(n+1)}{3},\end{aligned}$$

and inequality (1.8) is proved. ∎

The following corollary provides a counterpart of the generalised triangle inequality in normed spaces.

COROLLARY 1.3 (Dragomir, 2004 [**47**]). *Let* $(X, \|\cdot\|)$ *be a normed space and* $x_i \in X$, $p_i \geq 0$ $(i = 1, \ldots, n)$ *with* $\sum_{i=1}^{n} p_i = 1$*, then,*

$$0 \leq \sum_{i=1}^{n} p_i \|x_i\| - \left\| \sum_{i=1}^{n} p_i x_i \right\| \leq \max_{k=1,\ldots,n-1} \|\Delta x_k\| \sum_{i,j=1}^{n} p_i p_j |j - i| \tag{1.9}$$

and in particular,

$$0 \leq \sum_{i=1}^{n} \|x_i\| - \left\| \sum_{i=1}^{n} x_i \right\| \leq \frac{n^2 - 1}{3} \max_{k=1,\ldots,n-1} \|\Delta x_k\| . \tag{1.10}$$

The proof is apparent from Theorem 1.1 on choosing $F : X \to \mathbb{R}$, $F(X) = \|x\|$ which is $L-$Lipschitzian with $L = 1$ since,

$$|\|x\| - \|y\|| \leq \|x - y\| \quad \text{for all } x, y \in X.$$

2. APPLICATIONS FOR f-DIVERGENCE

The following theorem holds [**47**].

THEOREM 2.1 (Dragomir, 2004 [**47**]). *Let* $f : \mathbb{R}_+ \to \mathbb{R}$ *be* $L-$*Lipschitzian on* $\mathbb{R}_+$*, then for all* $p, q \in \mathbb{R}_+^n$*, we have the inequality,*

$$\left| I_f(p, q) - Q_n f\left(\frac{P_n}{Q_n}\right) \right| \leq L \max_{k=1,\ldots,n-1} \left| \frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k} \right| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i - j| . \tag{2.1}$$

PROOF. We apply inequality (1.2) for $F = f$ and $p_i = \frac{q_i}{Q_n}$, $x_i = \frac{p_i}{q_i}$ to get,

$$\left| f\left(\frac{\sum_{i=1}^{n} p_i}{Q_n}\right) - \frac{1}{Q_n} \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \right| \leq L \max_{k=1,\ldots,n-1} \left| \frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k} \right| \frac{1}{Q_n^2} \sum_{i,j=1}^{n} q_i q_j |i - j| .$$

from where we obtain (2.1). ∎

COROLLARY 2.2 (Dragomir, 2004 [**47**]). *Let* $f : \mathbb{R}_+ \to \mathbb{R}$ *be* $L-$*Lipschitzian and normalised, then for any* $p, q \in \mathbb{R}_+^n$ *with* $P_n = Q_n$*, we have,*

$$0 \leq |I_f(p, q)| \leq L \max_{k=1,\ldots,n-1} \left| \frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k} \right| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i - j| . \tag{2.2}$$

COROLLARY 2.3 (Dragomir, 2004 [**47**]). *Let* $f : \mathbb{R}_+ \to \mathbb{R}$ *be differentiable convex, with a bounded derivative, that is,* $\|f'\|_\infty := \sup_{t \geq 0} |f'(t)| < \infty$*, then,*

$$0 \leq I_f(p, q) - Q_n f\left(\frac{P_n}{Q_n}\right) \leq \|f'\|_\infty \max_{k=1,\ldots,n-1} \left| \frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k} \right| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i - j| . \tag{2.3}$$

Moreover, if f is normalised and $P_n = Q_n$, then,

$$0 \leq I_f(p,q) \leq \|f'\|_\infty \max_{k=1,\dots,n-1} \left| \frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k} \right| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j| . \tag{2.4}$$

We note that, for the convex mapping $f(t) := -\log(t)$, $t > 0$,

$$I_f(p,q) = \sum_{i=1}^{n} q_i \left[-\log\left(\frac{p_i}{q_i}\right)\right] = \sum_{i=1}^{n} q_i \log\left(\frac{q_i}{p_i}\right) = KL(q,p) . \tag{2.5}$$

PROPOSITION 2.4. *If $p, q \in \mathbb{R}_+^n$ satisfies the condition,*

$$0 < m \leq r_k := \frac{p_k}{q_k} \quad \textit{for all } k = 1, \dots, n, \tag{2.6}$$

then,

$$\begin{aligned} 0 &\leq KL(q,p) - Q_n \log\left(\frac{Q_n}{P_n}\right) \\ &\leq \frac{1}{m} \max_{k=1,\dots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j| . \end{aligned} \tag{2.7}$$

PROOF. As $f(t) = -\log(t)$, then $f'(t) = -\frac{1}{t}$, $t > 0$. Clearly, f' in the interval $[m, \infty)$ is bounded and $\|f'\|_\infty = \sup_{t\in[m,\infty)} \left|\frac{1}{t}\right| = \frac{1}{m} < \infty$. Applying the inequality (2.3), we deduce (2.7). ∎

REMARK 2.1. If we assume that $P_n = Q_n$, then $m \leq 1$ and (2.7) becomes,

$$0 \leq KL(q,p) \leq \frac{1}{m} \max_{k=1,\dots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j| . \tag{2.8}$$

We also know that for $f(t) = t \log t$, $t > 0$, the f-divergence is $f(p,q) = KL(p,q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right)$.

The following proposition also holds.

PROPOSITION 2.5. *Let $p, q \in \mathbb{R}_+^n$ satisfy the condition*

$$0 < m \leq r_k \leq M < \infty \quad \textit{for all } k = 1, \dots, n. \tag{2.9}$$

We have then the inequality,

$$\begin{aligned} 0 &\leq KL(q,p) - P_n \log\left(\frac{P_n}{Q_n}\right) \\ &\leq \max\{|\log(Ml)|, |\log(ml)|\} \\ &\quad \times \max_{k=1,\dots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j| . \end{aligned} \tag{2.10}$$

PROOF. For the mapping $f(t) = t \log t$, $t > 0$, $f'(t) = \log(t) + 1$. On the interval $[m, M]$ we have,

$$\log(m) + 1 \leq f'(M) \leq \log(M) + 1, \quad t \in [m, M] .$$

Applying (2.3), we deduce (2.10). ∎

REMARK 2.2. If we assume that $P_n = Q_n$, then $m \le 1 \le M$ and (2.10) becomes,

$$0 \le KL(q,p) \le \max\{|\log(Ml)|, |\log(ml)|\} \times \max_{k=1,\ldots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j|. \tag{2.11}$$

Let $f(t) = (\sqrt{t} - 1)^2$, $t > 0$, then I_f gives the *Hellinger distance,*

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

Using Corollary 2.3, we can state the following.

PROPOSITION 2.6. *Let $p, q \in \mathbb{R}_+^n$ satisfy the condition (2.9), then we have the inequality,*

$$0 \le h^2(p,q) - \left(\sqrt{P_n} - \sqrt{Q_n}\right)^2 \le \max\left\{\frac{|\sqrt{m} - 1|}{\sqrt{m}}, \frac{\left|\sqrt{M} - 1\right|}{\sqrt{M}}\right\} \times \max_{k=1,\ldots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j|. \tag{2.12}$$

PROOF. As $f(t) = (\sqrt{t} - 1)^2$, $t > 0$, then $f'(t) = 1 - \frac{1}{\sqrt{t}}$. If we consider the mapping f' restricted to the interval $[m, M] \subset (0, \infty)$, then we observe that,

$$\frac{|\sqrt{m} - 1|}{\sqrt{m}} \le f'(t) \le \frac{\left|\sqrt{M} - 1\right|}{\sqrt{M}}, \quad t \in [m, M]$$

and thus

$$\|f'\|_\infty \le \max\left\{\frac{|\sqrt{m} - 1|}{\sqrt{m}}, \frac{\left|\sqrt{M} - 1\right|}{\sqrt{M}}\right\}.$$

∎

REMARK 2.3. If we assume that $P_n = Q_n$, then $m \le 1 \le M$ and (2.12) becomes,

$$0 \le h^2(p,q) \le \max\left\{\frac{|1 - \sqrt{m}|}{\sqrt{m}}, \frac{\left|\sqrt{M} - 1\right|}{\sqrt{M}}\right\} \times \max_{k=1,\ldots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j|. \tag{2.13}$$

Consider the mapping $f(t) = t^\alpha$, $\alpha > 1$, $t > 0$ and the $\alpha-$*order entropy of Rényi* $D_\alpha(p,q)$.

PROPOSITION 2.7. *Let $p, q \in \mathbb{R}_+^n$ be such that,*

$$0 < r_k \le M < \infty \quad \textit{for all } k = 1, \ldots, n, \tag{2.14}$$

then,

$$0 \le D_\alpha(p,q) - P_n^\alpha Q_n^{1-\alpha} \le \alpha M^{\alpha-1} \max_{k=1,\dots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j| . \tag{2.15}$$

In particular, if $P_n = Q_n$, then $M \ge 1$ and (2.15) becomes,

$$0 \le D_\alpha(p,q) - Q_n \le \alpha M^{\alpha-1} \max_{k=1,\dots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j| . \tag{2.16}$$

The proof is apparent by Corollary 2.3 applied for $f(t) = t^\alpha$.

Finally, if we consider the χ^2-*distance,* $D_{\chi^2}(p,q)$, obtained from the Csiszár f-divergence for $f(t) = (t-1)^2$, $t > 0$, we have the following.

PROPOSITION 2.8. *Let $p, q \in \mathbb{R}_+^n$ fulfill the properties of (2.9), then we have the reverse inequality,*

$$\begin{aligned} 0 &\le D_{\chi^2}(p,q) - \frac{1}{Q_n}(P_n - Q_n)^2 \\ &\le 2\max\{|m-1|, |M-1|\} \\ &\quad \times \max_{k=1,\dots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j| . \end{aligned} \tag{2.17}$$

In particular, if $P_n = Q_n$, then $m \le 1 \le M$ and (2.17) becomes,

$$\begin{aligned} 0 &\le D_{\chi^2}(p,q) \\ &\le 2\max\{1-m, M-1\} \\ &\quad \times \max_{k=1,\dots,n-1} |r_{k+1} - r_k| \frac{1}{Q_n} \sum_{i,j=1}^{n} q_i q_j |i-j| . \end{aligned} \tag{2.18}$$

3. THE CASE OF l_1-NORM

The following general result can be stated as well [**47**]:

THEOREM 3.1 (Dragomir, 2004 [**47**]). *Let X,Y be two normed linear spaces with the norms $\|\cdot\|$ and $|\cdot|$ respectively. If $F : X \to Y$ is L - Lipschitzian, then for all $x_i \in X$, $p_i \ge 0$ with $\sum_{i=1}^n p_i = 1$ $(i = 1, \dots, n)$, we have,*

$$\left| F\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{i=1}^{n} p_i F(x_i) \right| \le L \sum_{i=1}^{n} p_i (1-p_i) \sum_{k=1}^{n-1} \|\Delta x_k\| . \tag{3.1}$$

PROOF. We follow the proof in [**47**].

As F is $L-$Lipschitzian, we can state (see the proof of Theorem 1.1) that

$$\left| F\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{j=1}^{n} p_j F(x_j) \right| \le 2L \sum_{1 \le i < j \le n} p_i p_j \|x_j - x_i\| . \tag{3.2}$$

Now, observe that, for $i < j$, we have,

$$x_j - x_i = \sum_{k=i}^{j-1} \Delta x_k,$$

where $\Delta x_k := x_{k+1} - x_k$ is the forward difference.

As in the proof of Theorem 1.1, we have,

$$\begin{aligned}\sum_{1\le i<j\le n} p_i p_j \|x_j - x_i\| &= \sum_{1\le i<j\le n} p_i p_j \left\|\sum_{k=i}^{j-1} \Delta x_k\right\| \le \sum_{1\le i<j\le n} p_i p_j \sum_{k=i}^{j-1} \|\Delta x_k\| \\ &\le \sum_{1\le i<j\le n} p_i p_j \sum_{k=1}^{n-1} \|\Delta x_k\| . \end{aligned} \tag{3.3}$$

Putting

$$I = \sum_{1\le i<j\le n} p_i p_j,$$

we observe that,

$$1 = \sum_{i,j=1}^{n} p_i p_j = 2 \sum_{1\le i<j\le n} p_i p_j + \sum_{i=1}^{n} p_i^2 = 2I + \sum_{i=1}^{n} p_i^2$$

from which we deduce,

$$I = \frac{1}{2} \sum_{i=1}^{n} p_i (1 - p_i) .$$

Using (3.2) - (3.3) we obtain (1.2). ∎

COROLLARY 3.2 (Dragomir, 2004 [**47**]). *With the above assumptions for F and x_i $(i = 1, \dots, n)$,*

$$\left|F\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \frac{1}{n}\sum_{i=1}^{n} F(x_i)\right| \le L \cdot \frac{n-1}{n} \sum_{k=1}^{n-1} \|\Delta x_k\| . \tag{3.4}$$

The proof is obvious by the above theorem, choosing $p_i = \frac{1}{n}$ $(i = 1, \dots, n)$.

The following corollary provides a counterpart of the generalised triangle inequality.

COROLLARY 3.3 (Dragomir, 2004 [**47**]). *Let $(X, \|\cdot\|)$ be a normed space and $x_i \in X$, $p_i \ge 0$ $(i = 1, \dots, n)$ with $\sum_{i=1}^{n} p_i = 1$, then,*

$$0 \le \sum_{i=1}^{n} p_i \|x_i\| - \left\|\sum_{i=1}^{n} p_i x_i\right\| \le \sum_{i=1}^{n} p_i (1 - p_i) \sum_{k=1}^{n-1} \|\Delta x_k\| \tag{3.5}$$

and in particular

$$0 \le \sum_{i=1}^{n} \|x_i\| - \left\|\sum_{i=1}^{n} x_i\right\| \le (n-1) \sum_{k=1}^{n-1} \|\Delta x_k\| . \tag{3.6}$$

The proof is by Theorem 3.1 on choosing $F : X \to \mathbb{R}$, $F(X) = \|x\|$ which is $L-$Lipschitzian with $L = 1$ since

$$|\|x\| - \|y\|| \le \|x - y\| \quad \text{for all } x, y \in X.$$

4. Applications for f-Divergence

The following theorem holds [**47**].

THEOREM 4.1 (Dragomir, 2004 [**47**]). *Let $f:\mathbb{R}_{+}\to\mathbb{R}$ be $L-$Lipschitzian on $\mathbb{R}_{+}$, then for all $p,q\in\mathbb{R}_{+}^{n}$,*

$$\left|I_{f}\left(p,q\right)-Q_{n}f\left(\frac{P_{n}}{Q_{n}}\right)\right|\leq\frac{L}{Q_{n}}\sum_{i=1}^{n}q_{i}\left(Q_{n}-q_{i}\right)\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}}-\frac{p_{k}}{q_{k}}\right|. \tag{4.1}$$

PROOF. We apply inequality (3.1) for $F=f$ and $p_{i}=\frac{q_{i}}{Q_{n}}$, $x_{i}=\frac{p_{i}}{q_{i}}$ $(i=1,\ldots,n)$ to get,

$$\left|f\left(\frac{P_{n}}{Q_{n}}\right)-\frac{1}{Q_{n}}\sum_{i=1}^{n}q_{i}f\left(\frac{p_{i}}{q_{i}}\right)\right|\leq L\sum_{i=1}^{n}\frac{q_{i}\left(Q_{n}-q_{i}\right)}{Q_{n}^{2}}\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}}-\frac{p_{k}}{q_{k}}\right|,$$

from which we obtain (4.1). ∎

COROLLARY 4.2 (Dragomir, 2004 [**47**]). *Let $f:\mathbb{R}_{+}\to\mathbb{R}$ be $L-$Lipschitzian and normalised, then for any $p,q\in\mathbb{R}_{+}^{n}$ with $P_{n}=Q_{n}$, we have,*

$$0\leq\left|I_{f}\left(p,q\right)\right|\leq\frac{L}{Q_{n}}\sum_{i=1}^{n}q_{i}\left(Q_{n}-q_{i}\right)\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}}-\frac{p_{k}}{q_{k}}\right|. \tag{4.2}$$

COROLLARY 4.3 (Dragomir, 2004 [**47**]). *If $f:\mathbb{R}_{+}\to\mathbb{R}$ is differentiable convex, with a bounded derivative, then,*

$$\begin{aligned}0&\leq I_{f}\left(p,q\right)-Q_{n}f\left(\frac{P_{n}}{Q_{n}}\right)\\&\leq\frac{\left\|f'\right\|_{\infty}}{Q_{n}}\sum_{i=1}^{n}q_{i}\left(Q_{n}-q_{i}\right)\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}}-\frac{p_{k}}{q_{k}}\right|.\end{aligned} \tag{4.3}$$

Moreover, if f is normalised and $P_{n}=Q_{n}$, then,

$$0\leq I_{f}\left(p,q\right)\leq\frac{\left\|f'\right\|_{\infty}}{Q_{n}}\sum_{i=1}^{n}q_{i}\left(Q_{n}-q_{i}\right)\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}}-\frac{p_{k}}{q_{k}}\right|. \tag{4.4}$$

The following proposition for the Kullback-Leibler distance holds.

PROPOSITION 4.4. *Let $p,q\in\mathbb{R}_{+}^{n}$ satisfy the condition,*

$$0<m\leq r_{k}:=\frac{p_{k}}{q_{k}}\quad\textit{for all }k=1,\ldots,n, \tag{4.5}$$

then,

$$\begin{aligned}0&\leq KL\left(q,p\right)-Q_{n}\log\left(\frac{Q_{n}}{P_{n}}\right)\\&\leq\frac{1}{mQ_{n}}\sum_{i=1}^{n}q_{i}\left(Q_{n}-q_{i}\right)\sum_{k=1}^{n-1}\left|r_{k+1}-r_{k}\right|.\end{aligned} \tag{4.6}$$

PROOF. As $f\left(t\right)=-\log\left(t\right)$, then $f'\left(t\right)=-\frac{1}{t}$, $t>0$, f' in the interval $[m,\infty)$ is bounded and $\left\|f'\right\|_{\infty}=\sup_{t\in[m,\infty)}\left|\frac{1}{t}\right|=\frac{1}{m}<\infty$. Applying the inequality (4.3), we deduce (4.6). ∎

REMARK 4.1. If we assume that $P_n = Q_n$, then $m \leq 1$ and (4.6) becomes,

$$0 \leq KL(q,p) \leq \frac{1}{mQ_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k| . \tag{4.7}$$

We also know that for $f(t) = t \log t$, $t > 0$, the Csiszár f-divergence is,

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

The following proposition also holds.

PROPOSITION 4.5. *Let $p, q \in \mathbb{R}_+^n$ satisfy the condition,*

$$0 < m \leq r_k \leq M < \infty \quad \text{for all } k = 1, \ldots, n, \tag{4.8}$$

then,

$$\begin{aligned} 0 &\leq KL(q,p) - P_n \log \left(\frac{P_n}{Q_n} \right) \\ &\leq \frac{\max \{ |\log (Ml)|, |\log (ml)| \}}{Q_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k| . \end{aligned} \tag{4.9}$$

PROOF. For the mapping $f(t) = t \log t$, $t > 0$, we have $f'(t) = \log(t) + 1$. On the interval $[m, M]$ we have,

$$\log m + 1 \leq f'(t) \leq \log M + 1, \quad t \in [m, M]$$

and hence

$$|f'(t)| \leq \max \{ |\log (Ml)|, |\log (ml)| \}, \quad t \in [m, M].$$

Applying (4.3), we deduce (4.9). ∎

REMARK 4.2. If we assume that $P_n = Q_n$, then (2.10) becomes,

$$\begin{aligned} 0 &\leq KL(q,p) \\ &\leq \frac{\max \{ |\log (Ml)|, |\log (ml)| \}}{Q_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k| . \end{aligned} \tag{4.10}$$

Let $f(t) = \left(\sqrt{t} - 1\right)^2$, $t > 0$, then I_f gives the *Hellinger distance*

$$h^2(p,q) = \sum_{i=1}^{n} \left(\sqrt{p_i} - \sqrt{q_i}\right)^2,$$

Using Corollary 4.3, we may state the following proposition.

PROPOSITION 4.6. *Assume that $p, q \in \mathbb{R}_+^n$ satisfy the condition (4.8), then,*

$$\begin{aligned} 0 &\leq h^2(p,q) - \left(\sqrt{P_n} - \sqrt{Q_n}\right)^2 \\ &\leq \max \left\{ \frac{|\sqrt{m} - 1|}{\sqrt{m}}, \frac{\left|\sqrt{M} - 1\right|}{\sqrt{M}} \right\} \\ &\quad \times \frac{1}{Q_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k| . \end{aligned} \tag{4.11}$$

PROOF. As $f(t) = (\sqrt{t}-1)^2$, $t > 0$, then $f'(t) = 1 - \frac{1}{\sqrt{t}}$. If we consider the mapping f' restricted to the interval $[m, M] \subset (0, \infty)$, then we observe that,

$$\frac{|\sqrt{m}-1|}{\sqrt{m}} \leq f'(t) \leq \frac{\left|\sqrt{M}-1\right|}{\sqrt{M}}, \quad t \in [m, M]$$

and thus

$$\|f'\|_\infty \leq \max\left\{\frac{|\sqrt{m}-1|}{\sqrt{m}}, \frac{\left|\sqrt{M}-1\right|}{\sqrt{M}}\right\}.$$

∎

REMARK 4.3. If we assume that $P_n = Q_n$, then $m \leq 1 \leq M$ and (2.12) becomes,

$$0 \leq h^2(p,q) \leq \max\left\{\frac{1-\sqrt{m}}{\sqrt{m}}, \frac{\sqrt{M}-1}{\sqrt{M}}\right\} \frac{1}{Q_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k|. \tag{4.12}$$

Consider now the mapping $f(t) = t^\alpha$, $\alpha > 1$, $t > 0$.

PROPOSITION 4.7. *Let $p, q \in \mathbb{R}_+^n$ be such that,*

$$0 < r_k \leq M < \infty \quad \textit{for all } k = 1, \ldots, n, \tag{4.13}$$

then,

$$0 \leq D_\alpha(p,q) - P_n^\alpha Q_n^{1-\alpha} \leq \alpha M^{\alpha-1} \frac{1}{Q_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k|. \tag{4.14}$$

In particular, if $P_n = Q_n$, then $M \geq 1$ and (4.14) becomes,

$$0 \leq D_\alpha(p,q) - Q_n \leq \alpha M^{\alpha-1} \frac{1}{Q_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k|. \tag{4.15}$$

The proof is obvious by Corollary 4.3 applied for $f(t) = t^\alpha$.

Finally, if we consider the χ^2-*distance* obtained from the Csiszár f-divergence for $f(t) = (t-1)^2$, then we can state the following.

PROPOSITION 4.8. *Let $p, q \in \mathbb{R}_+^n$ fulfill the conditions of (2.9), then,*

$$0 \leq D_{\chi^2}(p,q) - \frac{1}{Q_n}(P_n - Q_n)^2 \leq 2\max\{|m-1|, |M-1|\} \frac{1}{Q_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k|. \tag{4.16}$$

In particular, if $P_n = Q_n$, then (4.16) becomes,

$$0 \leq D_{\chi^2}(p,q) \leq 2\max\{1-m, M-1\} \frac{1}{Q_n} \sum_{i=1}^{n} q_i (Q_n - q_i) \sum_{k=1}^{n-1} |r_{k+1} - r_k|. \tag{4.17}$$

5. THE CASE OF l_p-NORM

Finally, the following general result may be stated [**47**]:

THEOREM 5.1 (Dragomir, 2004 [**47**]). *Let $X, Y, F, (p_i)_{i=\overline{1,n}}$ be as in Theorem 3.1, then we have the inequality:*

$$\left| F\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{i=1}^{n} p_i F(x_i)\right| \le L \sum_{i,j=1}^{n} p_i p_j |j-i|^{\frac{1}{q}} \left(\sum_{k=1}^{n-1} \|\Delta x_k\|^p\right)^{\frac{1}{p}}, \tag{5.1}$$

where $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$.

PROOF. We follow the proof in [**47**].

As in the proof of Theorem 1.1, we have,

$$\left| F\left(\sum_{i=1}^{n} p_i x_i\right) - \sum_{j=1}^{n} p_j F(x_j)\right| \le 2L \sum_{1\le i<j\le n} p_i p_j \|x_j - x_i\|. \tag{5.2}$$

Also,

$$\sum_{1\le i<j\le n} p_i p_j \|x_j - x_i\| \le \sum_{1\le i<j\le n} p_i p_j \sum_{k=i}^{j-1} \|\Delta x_k\|. \tag{5.3}$$

Using Hölder's discrete inequality, we may write for $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$,

$$\begin{aligned}\sum_{k=i}^{j-1} \|\Delta x_k\| &\le \left(\sum_{k=i}^{j-1} 1\right)^{\frac{1}{q}} \left(\sum_{k=i}^{j-1} \|\Delta x_k\|^p\right)^{\frac{1}{p}} \\ &= (j-i)^{\frac{1}{q}} \left(\sum_{k=i}^{j-1} \|\Delta x_k\|^p\right)^{\frac{1}{p}} \\ &\le (j-i)^{\frac{1}{q}} \left(\sum_{k=1}^{n-1} \|\Delta x_k\|^p\right)^{\frac{1}{p}}\end{aligned}$$

and then, by (5.3),

$$\begin{aligned}\sum_{1\le i<j\le n} p_i p_j \|x_j - x_i\| &\le \sum_{1\le i<j\le n} p_i p_j (j-i)^{\frac{1}{q}} \left(\sum_{k=1}^{n-1} \|\Delta x_k\|^p\right)^{\frac{1}{p}} \\ &= \frac{1}{2} \sum_{i,j=1}^{n} p_i p_j |j-i|^{\frac{1}{q}} \left(\sum_{k=1}^{n-1} \|\Delta x_k\|^p\right)^{\frac{1}{p}}.\end{aligned}$$

Using (5.2) we deduce (5.1). ∎

COROLLARY 5.2 (Dragomir, 2004 [**47**]). *Let $(X, \|\cdot\|)$ be a normed space and $x_i \in X$, $p_i \ge 0$ $(i = 1, \dots, n)$ with $\sum_{i=1}^{n} p_i = 1$, then,*

$$0 \le \sum_{i=1}^{n} p_i \|x_i\| - \left\|\sum_{i=1}^{n} p_i x_i\right\| \le \sum_{i,j=1}^{n} p_i p_j |j-i|^{\frac{1}{q}} \left(\sum_{k=1}^{n-1} \|\Delta x_k\|^p\right)^{\frac{1}{p}}. \tag{5.4}$$

6. Applications for Csiszár f-Divergence

The following result for Csiszár f-divergence holds [**47**].

THEOREM 6.1 (Dragomir, 2004 [**47**]). *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be $L-$Lipschitzian on $\mathbb{R}_+$, then for all $p, q \in \mathbb{R}_+^n$, we have the inequality:*

$$\left| I_f(p,q) - Q_n f\left(\frac{P_n}{Q_n}\right)\right| \leq \frac{L}{Q_n}\sum_{i,j=1}^{n} q_i q_j |j-i|^{\frac{1}{q}} \left(\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k}\right|^p\right)^{\frac{1}{p}}, \tag{6.1}$$

where $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$.

PROOF. We apply (5.1) for $F = f$, $p_i = \frac{q_i}{Q_n}$, $x_i = \frac{p_i}{q_i}$ to get,

$$\left| f\left(\frac{P_n}{Q_n}\right) - \frac{1}{Q_n}\sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right)\right| \leq L\sum_{i,j=1}^{n} \frac{q_i}{Q_n}\cdot\frac{q_j}{Q_n} |j-i|^{\frac{1}{q}} \left(\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k}\right|^p\right)^{\frac{1}{p}},$$

from which we obtain (6.1). ∎

COROLLARY 6.2 (Dragomir, 2004 [**47**]). *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be $L-$Lipschitzian and normalised, then for all $p, q \in \mathbb{R}_+^n$ with $P_n = Q_n$, we have,*

$$0 \leq |I_f(p,q)| \leq \frac{c}{Q_n}\sum_{i,j=1}^{n} q_i q_j |j-i|^{\frac{1}{q}} \left(\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k}\right|^p\right)^{\frac{1}{p}}. \tag{6.2}$$

COROLLARY 6.3 (Dragomir, 2004 [**47**]). *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be differentiable convex with a bounded derivative, then,*

$$0 \leq I_f(p,q) - Q_n f\left(\frac{P_n}{Q_n}\right) \leq \frac{\|f'\|_\infty}{Q_n}\sum_{i,j=1}^{n} q_i q_j |j-i|^{\frac{1}{q}} \left(\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k}\right|^p\right)^{\frac{1}{p}}. \tag{6.3}$$

Moreover, if f is normalised and $P_n = Q_n$, then,

$$0 \leq I_f(p,q) \leq \frac{\|f'\|_\infty}{Q_n}\sum_{i,j=1}^{n} q_i q_j |j-i|^{\frac{1}{q}} \left(\sum_{k=1}^{n-1}\left|\frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k}\right|^p\right)^{\frac{1}{p}}. \tag{6.4}$$

REMARK 6.1. Further inequalities for particular divergences as in the previous two sections can be stated, but we omit the details.

CHAPTER 8

Reverses of Jensen's Inequality and f-Divergences

1. Introduction

If $x_i, y_i \in \mathbb{R}$ and $w_i \geq 0$ $(i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^{n} w_i = 1$ then we may consider the *Čebyšev functional*

$$T_w(x, y) := \sum_{i=1}^{n} w_i x_i y_i - \sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i y_i. \tag{1.1}$$

The following result is known in the literature as the *Grüss inequality*

$$|T_w(x, y)| \leq \frac{1}{4} (\Gamma - \gamma)(\Delta - \delta), \tag{1.2}$$

provided

$$-\infty < \gamma \leq x_i \leq \Gamma < \infty, \ \ -\infty < \delta \leq y_i \leq \Delta < \infty \tag{1.3}$$

for $i = 1, \ldots, n.$

The constant $\frac{1}{4}$ is sharp in the sense that it cannot be replaced by a smaller constant.

If we assume that $-\infty < \gamma \leq x_i \leq \Gamma < \infty$ for $i = 1, \ldots, n$, then by the Grüss inequality for $y_i = x_i$ and by the Schwarz's discrete inequality, we have

$$\sum_{i=1}^{n} w_i \left| x_i - \sum_{j=1}^{n} w_j x_j \right| \leq \left[\sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{j=1}^{n} w_j x_j \right)^2 \right]^{\frac{1}{2}} \leq \frac{1}{2} (\Gamma - \gamma). \tag{1.4}$$

In order to provide a reverse of the celebrated Jensen's inequality for convex functions, S.S. Dragomir obtained in 2002 [**45**] the following result:

THEOREM 1.1. *Let $f : [m, M] \to \mathbb{R}$ be a differentiable convex function on (m, M). If $x_i \in [m, M]$ and $w_i \geq 0$ $(i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^{n} w_i = 1$, then one has the counterpart of Jensen's weighted discrete inequality:*

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i \right) \\ &\leq \sum_{i=1}^{n} w_i f'(x_i) x_i - \sum_{i=1}^{n} w_i f'(x_i) \sum_{i=1}^{n} w_i x_i \\ &\leq \frac{1}{2} [f'(M) - f'(m)] \sum_{i=1}^{n} w_i \left| x_i - \sum_{j=1}^{n} w_j x_j \right|. \end{aligned} \tag{1.5}$$

REMARK 1.1. We notice that the inequality between the first and the second term in (1.5) was proved in 1994 by Dragomir & Ionescu, see [**62**].

On making use of (1.4), we can state the following string of reverse inequalities

$$
\begin{aligned}
0 &\leq \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i\right) \\
&\leq \sum_{i=1}^{n} w_i f'(x_i)\, x_i - \sum_{i=1}^{n} w_i f'(x_i) \sum_{i=1}^{n} w_i x_i \\
&\leq \frac{1}{2}\left[f'(M) - f'(m)\right] \sum_{i=1}^{n} w_i \left| x_i - \sum_{j=1}^{n} w_j x_j \right| \\
&\leq \frac{1}{2}\left[f'(M) - f'(m)\right] \left[\sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{j=1}^{n} w_j x_j\right)^2\right]^{\frac{1}{2}} \\
&\leq \frac{1}{4}\left[f'(M) - f'(m)\right](M - m),
\end{aligned} \tag{1.6}
$$

provided that $f : [m, M] \subset \mathbb{R} \to \mathbb{R}$ is a differentiable convex function on (m, M), $x_i \in [m, M]$ and $w_i \geq 0 \ \ (i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^{n} w_i = 1$.

REMARK 1.2. We notice that the inequality between the first, second and last term from (1.6) was proved in the general case of positive linear functionals in 2001 by S. S. Dragomir in [**39**].

2. REVERSE INEQUALITIES

The following reverse of the Jensen's inequality holds:

THEOREM 2.1 (Dragomir, 2013 [**59**]). *Let $f : I \to \mathbb{R}$ be a continuous convex function on the interval of real numbers I and $m, M \in \mathbb{R}$, $m < M$ with $[m, M] \subset \mathring{I}$, $\mathring{I}$ is the interior of I. If $x_i \in [m, M]$ and $w_i \geq 0 \ \ (i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^{n} w_i = 1$, then*

$$
\begin{aligned}
0 &\leq \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i\right) \\
&\leq \frac{\left(M - \sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i x_i - m\right)}{M - m} \Psi_f\left(\sum_{i=1}^{n} w_i x_i; m, M\right) \\
&\leq \frac{\left(M - \sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i x_i - m\right)}{M - m} \sup_{t \in (m, M)} \Psi_f(t; m, M) \\
&\leq \left(M - \sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i x_i - m\right) \frac{f'_-(M) - f'_+(m)}{M - m} \\
&\leq \frac{1}{4}(M - m)\left[f'_-(M) - f'_+(m)\right],
\end{aligned} \tag{2.1}
$$

where $\Psi_f(\cdot; m, M) : (m, M) \to \mathbb{R}$ is defined by

$$\Psi_f(t; m, M) = \frac{f(M) - f(t)}{M - t} - \frac{f(t) - f(m)}{t - m}.$$

We also have the inequality

$$0 \le \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i\right) \le \frac{\left(M - \sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i x_i - m\right)}{M-m} \Psi_f\left(\sum_{i=1}^{n} w_i x_i; m, M\right) \le \frac{1}{4}(M-m)\Psi_f\left(\sum_{i=1}^{n} w_i x_i; m, M\right) \le \frac{1}{4}(M-m)\sup_{t\in(m,M)} \Psi_f(t;m,M) \le \frac{1}{4}(M-m)\left[f'_-(M) - f'_+(m)\right], \tag{2.2}$$

provided that $\sum_{i=1}^{n} w_i x_i \in (m, M)$.

PROOF. By the convexity of f we have that

$$\begin{aligned} &\sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i\right) \\ &= \sum_{i=1}^{n} w_i f\left[\frac{m(M-x_i) + M(x_i - m)}{M-m}\right] - f\left(\sum_{i=1}^{n} w_i \left[\frac{m(M-x_i) + M(x_i - m)}{M-m}\right]\right) \\ &\le \sum_{i=1}^{n} w_i \frac{(M-x_i) f(m) + (x_i - m) f(M)}{M-m} - f\left(\frac{m\left(M - \sum_{i=1}^{n} w_i x_i\right) + M\left(\sum_{i=1}^{n} w_i x_i - m\right)}{M-m}\right) \\ &= \frac{\left(M - \sum_{i=1}^{n} w_i x_i\right) f(m) + \left(\sum_{i=1}^{n} w_i x_i - m\right) f(M)}{M-m} - f\left(\frac{m\left(M - \sum_{i=1}^{n} w_i x_i\right) + M\left(\sum_{i=1}^{n} w_i x_i - m\right)}{M-m}\right) := B. \end{aligned} \tag{2.3}$$

By denoting

$$\Delta_f(t;m,M) := \frac{(t-m) f(M) + (M-t) f(m)}{M-m} - f(t), \quad t \in [m, M]$$

we have

$$\begin{aligned} \Delta_f(t;m,M) &= \frac{(t-m) f(M) + (M-t) f(m) - (M-m) f(t)}{M-m} \\ &= \frac{(t-m) f(M) + (M-t) f(m) - (M-t+t-m) f(t)}{M-m} \\ &= \frac{(t-m)[f(M) - f(t)] - (M-t)[f(t) - f(m)]}{M-m} \\ &= \frac{(M-t)(t-m)}{M-m} \Psi_f(t;m,M) \end{aligned} \tag{2.4}$$

for any $t \in (m, M)$.

Therefore we have the equality

$$B = \frac{\left(M - \sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i x_i - m\right)}{M - m} \Psi_f\left(\sum_{i=1}^{n} w_i x_i; m, M\right) \tag{2.5}$$

provided that $\sum_{i=1}^{n} w_i x_i \in (m, M)$.

For $\sum_{i=1}^{n} w_i x_i = m$ or $\sum_{i=1}^{n} w_i x_i = M$ the inequality (2.1) is obvious. If $\sum_{i=1}^{n} w_i x_i \in (m, M)$, then

$$\begin{aligned}
\Psi_f\left(\sum_{i=1}^{n} w_i x_i; m, M\right) &\leq \sup_{t\in(m,M)} \Psi_f(t; m, M) \\
&= \sup_{t\in(m,M)} \left[\frac{f(M) - f(t)}{M - t} - \frac{f(t) - f(m)}{t - m}\right] \\
&\leq \sup_{t\in(m,M)} \left[\frac{f(M) - f(t)}{M - t}\right] + \sup_{t\in(m,M)} \left[-\frac{f(t) - f(m)}{t - m}\right] \\
&= \sup_{t\in(m,M)} \left[\frac{f(M) - f(t)}{M - t}\right] - \inf_{t\in(m,M)} \left[\frac{f(t) - f(m)}{t - m}\right] \\
&= f'_{-}(M) - f'_{+}(m)
\end{aligned}$$

which by (2.3) and (2.5) produces the desired result (2.1).

Since, obviously

$$\frac{\left(M - \sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i x_i - m\right)}{M - m} \leq \frac{1}{4}(M - m),$$

then by (2.3) and (2.5) we deduce the second inequality (2.2). The last part is clear. ∎

COROLLARY 2.2. *Let $f : I \to \mathbb{R}$ be a continuous convex function on the interval of real numbers I and $m, M \in \mathbb{R}$, $m < M$ with $[m, M] \subset \mathring{I}$. If $x_i \in [m, M]$, then we have the inequalities*

$$\begin{aligned}
0 &\leq \frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\
&\leq \frac{\left(M - \frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i - m\right)}{M - m} \Psi_f\left(\frac{1}{n}\sum_{i=1}^{n} x_i; m, M\right) \\
&\leq \frac{\left(M - \frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i - m\right)}{M - m} \sup_{t\in(m,M)} \Psi_f(t; m, M) \\
&\leq \left(M - \frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i - m\right) \frac{f'_{-}(M) - f'_{+}(m)}{M - m} \\
&\leq \frac{1}{4}(M - m)\left[f'_{-}(M) - f'_{+}(m)\right],
\end{aligned} \tag{2.6}$$

and

$$0 \leq \frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \tag{2.7}$$
$$\frac{\left(M - \frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} x_i - m\right)}{M-m}\Psi_f\left(\frac{1}{n}\sum_{i=1}^{n} x_i; m, M\right)$$
$$\leq \frac{1}{4}(M-m)\Psi_f\left(\frac{1}{n}\sum_{i=1}^{n} x_i; m, M\right)$$
$$\leq \frac{1}{4}(M-m)\sup_{t\in(m,M)}\Psi_f(t;m,M)$$
$$\leq \frac{1}{4}(M-m)\left[f'_-(M) - f'_+(m)\right],$$

where $\frac{1}{n}\sum_{i=1}^{n} x_i \in (m, M)$.

REMARK 2.1. Define the weighted arithmetic mean of the positive n-tuple $x = (x_1, ..., x_n)$ with the nonnegative weights $w = (w_1, ..., w_n)$ by

$$A_n(w,x) := \frac{1}{W_n}\sum_{i=1}^{n} w_i x_i$$

where $W_n := \sum_{i=1}^{n} w_i > 0$ and the weighted geometric mean of the same n-tuple, by

$$G_n(w,x) := \left(\prod_{i=1}^{n} x_i^{w_i}\right)^{1/W_n}.$$

It is well know that the following arithmetic mean-geometric mean inequality holds true

$$A_n(w,x) \geq G_n(w,x).$$

Applying the inequality between the first and third term in (2.6) for the convex function $f(t) = -\ln t, t > 0$ we have

$$1 \leq \frac{A_n(w,x)}{G_n(w,x)} \leq \exp\left[\frac{1}{Mm}(M - A_n(w,x))(A_n(w,x) - m)\right] \tag{2.8}$$
$$\leq \exp\left[\frac{1}{4}\frac{(M-m)^2}{mM}\right],$$

provided that $0 < m \leq x_i \leq M < \infty$ for $i \in \{1, ..., n\}$.

Also, if we apply the inequality (2.7) for the same function f we get that

$$1 \leq \frac{A_n(w,x)}{G_n(w,x)} \tag{2.9}$$
$$\leq \left[\left(\frac{M}{A_n(w,x)}\right)^{M-A_n(w,x)}\left(\frac{m}{A_n(w,x)}\right)^{A_n(w,x)-m}\right]^{-\frac{1}{4}(M-m)}$$
$$\leq \exp\left[\frac{1}{4}\frac{(M-m)^2}{mM}\right].$$

The following result also holds:

THEOREM 2.3 (Dragomir, 2013 [**59**]). *With the assumptions of Theorem 2.1, we have the inequalities*

$$0 \leq \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i\right) \leq 2\max\left\{\frac{M-\sum_{i=1}^{n} w_i x_i}{M-m}, \frac{\sum_{i=1}^{n} w_i x_i - m}{M-m}\right\} \times \left[\frac{f(m)+f(M)}{2} - f\left(\frac{m+M}{2}\right)\right] \leq \frac{1}{2}\max\left\{M-\sum_{i=1}^{n} w_i x_i, \sum_{i=1}^{n} w_i x_i - m\right\}\left[f'_-(M) - f'_+(m)\right]. \tag{2.10}$$

PROOF. First of all, we recall the following result obtained by the author in [**50**] that provides a refinement and a reverse for the weighted Jensen's discrete inequality:

$$n \min_{i\in\{1,\dots,n\}}\{p_i\}\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\right] \leq \frac{1}{P_n}\sum_{i=1}^{n} p_i f(x_i) - f\left(\frac{1}{P_n}\sum_{i=1}^{n} p_i x_i\right) n \max_{i\in\{1,\dots,n\}}\{p_i\}\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\right], \tag{2.11}$$

where $f : C \to \mathbb{R}$ is a convex function defined on the convex subset C of the linear space X, $\{x_i\}_{i\in\{1,\dots,n\}} \subset C$ are vectors and $\{p_i\}_{i\in\{1,\dots,n\}}$ are nonnegative numbers with $P_n := \sum_{i=1}^{n} p_i > 0$.

For $n = 2$ we deduce from (2.11) that

$$2\min\{t, 1-t\}\left[\frac{f(x)+f(y)}{2} - f\left(\frac{x+y}{2}\right)\right] \leq tf(x) + (1-t)f(y) - f(tx+(1-t)y) \leq 2\max\{t, 1-t\}\left[\frac{f(x)+f(y)}{2} - f\left(\frac{x+y}{2}\right)\right] \tag{2.12}$$

for any $x, y \in C$ and $t \in [0,1]$.

If we use the second inequality in (2.12) for the convex function $f : I \to \mathbb{R}$ and $m, M \in \mathbb{R}$, $m < M$ with $[m, M] \subset \mathring{I}$, we have for $t = \frac{M-\sum_{i=1}^{n} w_i x_i}{M-m}$ that

$$\frac{\left(M-\sum_{i=1}^{n} w_i x_i\right) f(m) + \left(\sum_{i=1}^{n} w_i x_i - m\right) f(M)}{M-m} - f\left(\frac{m\left(M-\sum_{i=1}^{n} w_i x_i\right) + M\left(\sum_{i=1}^{n} w_i x_i - m\right)}{M-m}\right) \leq 2\max\left\{\frac{M-\sum_{i=1}^{n} w_i x_i}{M-m}, \frac{\sum_{i=1}^{n} w_i x_i - m}{M-m}\right\} \times \left[\frac{f(m)+f(M)}{2} - f\left(\frac{m+M}{2}\right)\right]. \tag{2.13}$$

Utilizing the inequality (2.3) and (2.13) we deduce the first inequality in (2.10).

Since

$$\frac{\frac{f(m)+f(M)}{2}-f\left(\frac{m+M}{2}\right)}{M-m} = \frac{1}{4}\left[\frac{f(M)-f\left(\frac{m+M}{2}\right)}{M-\frac{m+M}{2}}-\frac{f\left(\frac{m+M}{2}\right)-f(m)}{\frac{m+M}{2}-m}\right]$$

and, by the gradient inequality, we have that

$$\frac{f(M)-f\left(\frac{m+M}{2}\right)}{M-\frac{m+M}{2}} \le f'_-(M)$$

and

$$\frac{f\left(\frac{m+M}{2}\right)-f(m)}{\frac{m+M}{2}-m} \ge f'_+(m),$$

then we get

$$\frac{\frac{f(m)+f(M)}{2}-f\left(\frac{m+M}{2}\right)}{M-m} \le \frac{1}{4}\left[f'_-(M)-f'_+(m)\right]. \tag{2.14}$$

On making use of (2.13) and (2.14) we deduce the last part of (2.10). ∎

COROLLARY 2.4. *With the assumptions in Corollary 2.2, we have the inequalities*

$$0 \le \frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \tag{2.15}$$

$$\le 2\max\left\{\frac{M-\frac{1}{n}\sum_{i=1}^{n}x_i}{M-m}, \frac{\frac{1}{n}\sum_{i=1}^{n}x_i - m}{M-m}\right\} \times\left[\frac{f(m)+f(M)}{2}-f\left(\frac{m+M}{2}\right)\right] \tag{2.16}$$

$$\le \frac{1}{2}\max\left\{M-\frac{1}{n}\sum_{i=1}^{n}x_i, \frac{1}{n}\sum_{i=1}^{n}x_i - m\right\}\left[f'_-(M)-f'_+(m)\right].$$

REMARK 2.2. Since, obviously,

$$\frac{M-\sum_{i=1}^{n}w_i x_i}{M-m}, \frac{\sum_{i=1}^{n}w_i x_i - m}{M-m} \le 1,$$

then we obtain from the first inequality in (2.10) the simpler, however coarser inequality, namely

$$0 \le \sum_{i=1}^{n} w_i f(x_i) - f\left(\sum_{i=1}^{n} w_i x_i\right) \le 2\left[\frac{f(m)+f(M)}{2}-f\left(\frac{m+M}{2}\right)\right]. \tag{2.17}$$

This inequality was obtained in 2008 by S. Simic in [**122**].

REMARK 2.3. With the assumptions in Remark 2.1 we have the following reverse of the arithmetic mean-geometric mean inequality

$$1 \le \frac{A_n(w,x)}{G_n(w,x)} \le \left(\frac{A(m,M)}{G(m,M)}\right)^{2\max\left\{\frac{M-A_n(w,x)}{M-m}, \frac{A_n(w,x)-m}{M-m}\right\}}, \tag{2.18}$$

where $A(m,M)$ is the arithmetic mean while $G(m,M)$ is the geometric mean of the positive numbers m and M.

3. Applications for the Hölder Inequality

If $x_i, y_i \geq 0$ for $i \in \{1, ..., n\}$, then the *Hölder inequality* holds true

$$\sum_{i=1}^{n} x_i y_i \leq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1/q},$$

where $p > 1, \frac{1}{p} + \frac{1}{q} = 1.$

Assume that $p > 1$. If $z_i \in \mathbb{R}$ for $i \in \{1, ..., n\}$, satisfies the bounds

$$0 < m \leq z_i \leq M < \infty \text{ for } i \in \{1, ..., n\}$$

and $w_i \geq 0 \ \ (i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^{n} w_i > 0$, then from (2.1) we have

$$\begin{aligned} 0 &\leq \frac{\sum_{i=1}^{n} w_i z_i^p}{W_n} - \left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n}\right)^p \\ &\leq \frac{\left(M - \frac{\sum_{i=1}^{n} w_i z_i}{W_n}\right)\left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n} - m\right)}{M - m} B_p(m, M) \\ &\leq p\frac{M^{p-1} - m^{p-1}}{M - m}\left(M - \frac{\sum_{i=1}^{n} w_i z_i}{W_n}\right)\left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n} - m\right) \\ &\leq \frac{1}{4} p(M - m)\left(M^{p-1} - m^{p-1}\right), \end{aligned} \tag{3.1}$$

where $\Psi_p(\cdot; m, M) : (m, M) \to \mathbb{R}$ is defined by

$$\Psi_p(t; m, M) = \frac{M^p - t^p}{M - t} - \frac{t^p - m^p}{t - m}$$

while

$$B_p(m, M) := \sup_{t \in (m, M)} \Psi_p(t; m, M). \tag{3.2}$$

From (2.2) we also have the inequality

$$\begin{aligned} 0 &\leq \frac{\sum_{i=1}^{n} w_i z_i^p}{W_n} - \left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n}\right)^p \\ &\leq \frac{1}{4}(M - m)\Psi_p\left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n}; m, M\right) \leq \frac{1}{4} p(M - m)\left(M^{p-1} - m^{p-1}\right). \end{aligned} \tag{3.3}$$

Proposition 3.1 (Dragomir, 2013 [**59**]). *If* $x_i \geq 0, y_i > 0$ *for* $i \in \{1, ..., n\}$, $p > 1, \frac{1}{p} + \frac{1}{q} = 1$ *and there exists the constants* $\gamma, \Gamma > 0$ *and such that*

$$\gamma \leq \frac{x_i}{y_i^{q-1}} \leq \Gamma \text{ for } i \in \{1, ..., n\},$$

then we have

$$\begin{aligned} 0 &\leq \frac{\sum_{i=1}^{n} x_i^p}{\sum_{i=1}^{n} y_i^q} - \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right)^p \\ &\leq \frac{B_p(\gamma, \Gamma)}{\Gamma - \gamma}\left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right)\left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right) \\ &\leq p\frac{\Gamma^{p-1} - \gamma^{p-1}}{\Gamma - \gamma}\left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right)\left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right) \\ &\leq \frac{1}{4} p(\Gamma - \gamma)\left(\Gamma^{p-1} - \gamma^{p-1}\right), \end{aligned} \tag{3.4}$$

and

$$0 \le \frac{\sum_{i=1}^n x_i^p}{\sum_{i=1}^n y_i^q} - \left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q}\right)^p \le \frac{1}{4}(\Gamma-\gamma)\Psi_p\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q};\gamma,\Gamma\right) \le \frac{1}{4}p(\Gamma-\gamma)\left(\Gamma^{p-1}-\gamma^{p-1}\right), \tag{3.5}$$

where $B_p(\cdot,\cdot)$ *and* $\Psi_p(\cdot;\cdot,\cdot)$ *are defined above.*

PROOF. The inequalities (3.4) and (3.5) follow from (3.1) and (3.3) by choosing

$$z_i = \frac{x_i}{y_i^{q-1}} \text{ and } w_i = y_i^q.$$

The details are omitted. ∎

REMARK 3.1. We observe that for $p = q = 2$ we have $\Psi_2(t;\gamma,\Gamma) = \Gamma - \gamma = B_2(\gamma,\Gamma)$ and then from the first inequality in (3.4) we get the following reverse of the Cauchy-Bunyakovsky-Schwarz inequality:

$$\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n x_i y_i\right)^2 \le \left(\Gamma - \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^2}\right)\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^2} - \gamma\right)\left(\sum_{i=1}^n y_i^2\right)^2 \tag{3.6}$$

provided that $x_i \ge 0$, $y_i > 0$ for $i \in \{1,...,n\}$ and there exists the constants $\gamma, \Gamma > 0$ such that

$$\gamma \le \frac{x_i}{y_i} \le \Gamma \ for\ i \in \{1,...,n\}.$$

COROLLARY 3.2 (Dragomir, 2013 [**59**]). *With the assumptions of Proposition 3.1 we have the following additive reverses of the Hölder inequality*

$$\begin{aligned} 0 &\le \left(\sum_{i=1}^n x_i^p\right)^{1/p}\left(\sum_{i=1}^n y_i^q\right)^{1/q} - \sum_{i=1}^n x_i y_i \\ &\le \left[\frac{B_p(\gamma,\Gamma)}{\Gamma-\gamma}\right]^{1/p}\left(\Gamma - \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q}\right)^{1/p}\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q} - \gamma\right)^{1/p} \\ &\times \sum_{i=1}^n y_i^q \\ &\le p^{1/p}\left(\frac{\Gamma^{p-1}-\gamma^{p-1}}{\Gamma-\gamma}\right)^{1/p}\left(\Gamma - \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q}\right)^{1/p}\left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q} - \gamma\right)^{1/p} \\ &\times \sum_{i=1}^n y_i^q \\ &\le \frac{1}{4^{1/p}}p^{1/p}(\Gamma-\gamma)^{1/p}\left(\Gamma^{p-1}-\gamma^{p-1}\right)^{1/p}\sum_{i=1}^n y_i^q \end{aligned} \tag{3.7}$$

and

(3.8)
$$0 \leq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1/q} - \sum_{i=1}^{n} x_i y_i$$
$$\leq \frac{1}{4^{1/p}} (\Gamma - \gamma)^{1/p} \Psi_p^{1/p} \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}; m, M\right) \sum_{i=1}^{n} y_i^q$$
$$\leq \frac{1}{4^{1/p}} p^{1/p} (\Gamma - \gamma)^{1/p} \left(\Gamma^{p-1} - \gamma^{p-1}\right)^{1/p} \sum_{i=1}^{n} y_i^q$$

where $p > 1$ *and* $\frac{1}{p} + \frac{1}{q} = 1$.

PROOF. By multiplying in (3.4) with $(\sum_{i=1}^{n} y_i^q)^p$ we have

$$\sum_{i=1}^{n} x_i^p \left(\sum_{i=1}^{n} y_i^q\right)^{p-1} - \left(\sum_{i=1}^{n} x_i y_i\right)^p$$
$$\leq \frac{B_p(\gamma, \Gamma)}{\Gamma - \gamma} \left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right) \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right) \left(\sum_{i=1}^{n} y_i^q\right)^p$$
$$\leq p \frac{\Gamma^{p-1} - \gamma^{p-1}}{\Gamma - \gamma} \left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right) \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right) \left(\sum_{i=1}^{n} y_i^q\right)^p$$
$$\leq \frac{1}{4} p (\Gamma - \gamma) \left(\Gamma^{p-1} - \gamma^{p-1}\right) \left(\sum_{i=1}^{n} y_i^q\right)^p,$$

which is equivalent to

(3.9)
$$\sum_{i=1}^{n} x_i^p \left(\sum_{i=1}^{n} y_i^q\right)^{p-1}$$
$$\leq \left(\sum_{i=1}^{n} x_i y_i\right)^p + \frac{B_p(\gamma, \Gamma)}{\Gamma - \gamma} \left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right) \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right)$$
$$\times \left(\sum_{i=1}^{n} y_i^q\right)^p$$
$$\leq \left(\sum_{i=1}^{n} x_i y_i\right)^p + p \left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right) \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right)$$
$$\times \left(\sum_{i=1}^{n} y_i^q\right)^p \frac{\Gamma^{p-1} - \gamma^{p-1}}{\Gamma - \gamma}$$
$$\leq \left(\sum_{i=1}^{n} x_i y_i\right)^p + \frac{1}{4} p (\Gamma - \gamma) \left(\Gamma^{p-1} - \gamma^{p-1}\right) \left(\sum_{i=1}^{n} y_i^q\right)^p.$$

Taking the power $1/p$ with $p > 1$ and employing the following elementary inequality that state that for $p > 1$ and $\alpha, \beta > 0$,

$$(\alpha + \beta)^{1/p} \leq \alpha^{1/p} + \beta^{1/p}$$

we have from the first part of (3.9) that

$$\left(\sum_{i=1}^{n} x_i y_i\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1-\frac{1}{p}} \tag{3.10}$$
$$\le \sum_{i=1}^{n} x_i y_i + \left[\frac{B_p(\gamma,\Gamma)}{\Gamma-\gamma}\right]^{1/p} \left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right)^{1/p} \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right)^{1/p}$$
$$\times \sum_{i=1}^{n} y_i^q$$

and since $1 - \frac{1}{p} = \frac{1}{q}$ we get from (3.10) the first inequality in (3.7). The rest is obvious.

The inequality (3.8) can be proved in a similar manner, however the details are omitted. ∎

If $z_i \in \mathbb{R}$ for $i \in \{1, ..., n\}$, satisfies the bounds

$$0 < m \le z_i \le M < \infty \text{ for } i \in \{1, ..., n\}$$

and $w_i \ge 0 \ \ (i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^{n} w_i > 0$, then from (2.10) we also have the inequality

$$0 \le \frac{\sum_{i=1}^{n} w_i z_i^p}{W_n} - \left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n}\right)^p \tag{3.11}$$
$$\le 2\left[\frac{m^p + M^p}{2} - \left(\frac{m+M}{2}\right)^p\right]$$
$$\times \max\left\{\frac{M - \frac{\sum_{i=1}^{n} w_i z_i}{W_n}}{M-m}, \frac{\frac{\sum_{i=1}^{n} w_i z_i}{W_n} - m}{M-m}\right\}$$
$$\le \frac{1}{2} p\left(M^{p-1} - m^{p-1}\right) \max\left\{M - \frac{\sum_{i=1}^{n} w_i z_i}{W_n}, \frac{\sum_{i=1}^{n} w_i z_i}{W_n} - m\right\}.$$

From the inequality (3.11) we can state:

PROPOSITION 3.3 (Dragomir, 2013 [**59**]). *With the assumptions of Proposition 3.1 we have*

$$0 \le \frac{\sum_{i=1}^{n} x_i^p}{\sum_{i=1}^{n} y_i^q} - \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right)^p \tag{3.12}$$
$$\le 2 \cdot \frac{\frac{\gamma^p + \Gamma^p}{2} - \left(\frac{\gamma+\Gamma}{2}\right)^p}{\Gamma - \gamma} \max\left\{\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}, \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right\}$$
$$\le \frac{1}{2} p\left(\Gamma^{p-1} - \gamma^{p-1}\right) \max\left\{\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}, \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right\}.$$

Finally, the following additive reverse of the Hölder inequality can be stated as well:

COROLLARY 3.4 (Dragomir, 2013 [**59**]). *With the assumptions of Proposition 3.1 we have*

$$0 \leq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1/q} - \sum_{i=1}^{n} x_i y_i \tag{3.13}$$
$$\leq 2^{1/p} \cdot \left(\frac{\frac{\gamma^p+\Gamma^p}{2} - \left(\frac{\gamma+\Gamma}{2}\right)^p}{\Gamma - \gamma}\right)^{1/p}$$
$$\times \max\left\{\left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right)^{1/p}, \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right)^{1/p}\right\} \sum_{i=1}^{n} y_i^q$$
$$\leq \frac{1}{2^{1/p}} p^{1/p} \max\left\{\left(\Gamma - \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right)^{1/p}, \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q} - \gamma\right)^{1/p}\right\}$$
$$\times \left(\Gamma^{p-1} - \gamma^{p-1}\right)^{1/p} \sum_{i=1}^{n} y_i^q.$$

REMARK 3.2. As a simpler, however coarser inequality we have the following result:

$$0 \leq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1/q} - \sum_{i=1}^{n} x_i y_i \tag{3.14}$$
$$\leq 2^{1/p} \cdot \left[\frac{\gamma^p + \Gamma^p}{2} - \left(\frac{\gamma + \Gamma}{2}\right)^p\right]^{1/p} \sum_{i=1}^{n} y_i^q,$$

where x_i and y_i are as above.

4. APPLICATIONS FOR f-DIVERGENCE

Consider the f-divergence

$$I_f(p,q) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \tag{4.1}$$

defined on the set of probability distributions $p, q \in \mathbb{P}^n$, where f is convex on $(0, \infty)$. It is assumed that $f(u)$ is zero and strictly convex at $u = 1$. By appropriately defining this convex function, various divergences are derived.

The following result holds:

PROPOSITION 4.1 (Dragomir, 2013 [**59**]). *Let $f : (0,\infty) \to \mathbb{R}$ be a convex function with the property that $f(1) = 0$. Assume that $p, q \in \mathbb{P}^n$ and there exists the constants $0 < r < 1 < R < \infty$ such that*

$$r \leq \frac{p_i}{q_i} \leq R \text{ for } i \in \{1, ..., n\}. \tag{4.2}$$

Then we have the inequalities

$$0 \leq I_f(p,q) \leq \frac{(R-1)(1-r)}{R-r} \sup_{t\in(r,R)} \Psi_f(t; r, R) \tag{4.3}$$
$$\leq (R-1)(1-r) \frac{f'_-(R) - f'_+(r)}{R-r}$$
$$\leq \frac{1}{4}(R-r)\left[f'_-(R) - f'_+(r)\right],$$

and $\Psi_f(\cdot; r, R) : (r, R) \to \mathbb{R}$ *is defined by*

$$\Psi_f(t; r, R) = \frac{f(R) - f(t)}{R - t} - \frac{f(t) - f(r)}{t - r}.$$

We also have the inequality

$$I_f(p, q) \le \frac{1}{4}(R - r)\frac{f(R)(1 - r) + f(r)(R - 1)}{(R - 1)(1 - r)} \le \frac{1}{4}(R - r)\left[f'_-(R) - f'_+(r)\right]. \tag{4.4}$$

The proof follows by Theorem 2.1 by choosing $w_i = q_i$, $x_i = \frac{p_i}{q_i}$, $m = r$ and $M = R$ and performing the required calculations. The details are omitted.

Utilising the same approach and Theorem 2.3 we can also state that:

PROPOSITION 4.2 (Dragomir, 2013 [**59**]). *With the assumptions of Proposition 4.1 we have*

$$0 \le I_f(p, q) \le 2\max\left\{\frac{R - 1}{R - r}, \frac{1 - r}{R - r}\right\} \times \left[\frac{f(r) + f(R)}{2} - f\left(\frac{r + R}{2}\right)\right] \le \frac{1}{2}\max\{R - 1, 1 - r\}\left[f'_-(R) - f'_+(r)\right]. \tag{4.5}$$

The above results can be utilized to obtain various inequalities for the divergence measures in Information Theory that are particular instances of f-divergence.

Consider the Kullback-Leibler divergence

$$KL(p, q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right),\ p, q \in \mathbb{P}^n.$$

For the convex function $f : (0, \infty) \to \mathbb{R}$, $f(t) = -\ln t$ we have

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

If $p, q \in \mathbb{P}^n$ such that there exists the constants $0 < r < 1 < R < \infty$ with

$$r \le \frac{p_i}{q_i} \le R \text{ for } i \in \{1, ..., n\}, \tag{4.6}$$

then we get from the second inequality in (4.3) that

$$0 \le KL(q, p) \le \frac{(R - 1)(1 - r)}{rR}, \tag{4.7}$$

from the first inequality in (4.4) that

$$0 \le KL(q, p) \le \frac{1}{4}(R - r)\ln\left[R^{-\frac{1}{R-1}} r^{-\frac{1}{1-r}}\right]$$

and from the first inequality in (4.5) that

$$0 \le KL(q, p) \le 2\max\left\{\frac{R - 1}{R - r}, \frac{1 - r}{R - r}\right\}\ln\left(\frac{A(r, R)}{G(r, R)}\right) \tag{4.8}$$

where $A(r, R)$ is the arithmetic mean and $G(r, R)$ is the geometric mean of the positive numbers r and R.

For the convex function $f:(0,\infty)\to\mathbb{R}$, $f(t)=t\ln t$ we have

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

If $p,q\in\mathbb{P}^n$ such that there exists the constants $0<r<1<R<\infty$ with the property (4.6), then we get from the second inequality in (4.3) that

$$0\le KL(p,q)\le\frac{(R-1)(1-r)}{L(r,R)}, \tag{4.9}$$

where $L(r,R)$ is the Logarithmic mean of r, R, namely

$$L(r,R)=\frac{R-r}{\ln R-\ln r}.$$

From the first inequality in (4.4) we also have:

$$0\le KL(p,q)\le\frac{1}{4}(R-r)\frac{R-r+\ln\left(R^{1-r}r^{R-1}\right)}{(R-1)(1-r)}. \tag{4.10}$$

Finally, by the first inequality in (4.5) we have

$$0\le KL(p,q)\le 2\max\left\{\frac{R-1}{R-r},\frac{1-r}{R-r}\right\}\ln\left[\frac{G\left(r^r,R^R\right)}{[A(r,R)]^{A(r,R)}}\right]. \tag{4.11}$$

5. MORE REVERSE INEQUALITIES

For the *Lebesgue measurable* function $g:[\alpha,\beta]\to\mathbb{R}$ we introduce the *Lebesgue p-norms* defined as

$$\|g\|_{[\alpha,\beta],p}:=\left(\int_\alpha^\beta|g(t)|^p\,dt\right)^{1/p}\text{ if }g\in L_p[\alpha,\beta],$$

for $p\ge 1$ and

$$\|g\|_{[\alpha,\beta],\infty}:=ess\sup_{t\in[\alpha,\beta]}|g(t)|\text{ if }g\in L_\infty[\alpha,\beta],$$

for $p=\infty$.

The following result also holds:

THEOREM 5.1 (Dragomir, 2013 [**60**]). *Let $\Phi:I\to\mathbb{R}$ be a continuous convex function on the interval of real numbers I and $m,M\in\mathbb{R}$, $m<M$ with $[m,M]\subset\mathring{I}$, $\mathring{I}$ is the interior of I. If $x_i\in I$ and $w_i\ge 0$ for $i\in\{1,...,n\}$ with $\sum_{i=1}^n w_i=1$, denote $\bar{x}_w:=\sum_{i=1}^n w_i x_i\in I$, then we have the inequality*

$$\begin{aligned}0&\le\sum_{i=1}^{n}w_i\Phi(x_i)-\Phi(\bar{x}_w)\\&\le\frac{(M-\bar{x}_w)\int_m^{\bar{x}_w}|\Phi'(t)|\,dt+(\bar{x}_w-m)\int_{\bar{x}_w}^M|\Phi'(t)|\,dt}{M-m}:=\Theta_\Phi(\bar{x}_w;m,M),\end{aligned} \tag{5.1}$$

where $\Theta_\Phi(\bar{x}_w;m,M)$ satisfies the bounds

$$\Theta_\Phi(\bar{x}_w;m,M)\le\begin{cases}\left[\frac{1}{2}+\frac{\left|\bar{x}_w-\frac{m+M}{2}\right|}{M-m}\right]\int_m^M|\Phi'(t)|\,dt\\[2ex]\left[\frac{1}{2}\int_m^M|\Phi'(t)|\,dt+\frac{1}{2}\left|\int_{\bar{x}_w}^M|\Phi'(t)|\,dt-\int_m^{\bar{x}_w}|\Phi'(t)|\,dt\right|\right],\end{cases} \tag{5.2}$$

$$\begin{aligned}
(5.3)\qquad &\Theta_{\Phi}\left(\bar{x}_w; m, M\right) \\
&\leq \frac{\left(\bar{x}_w-m\right)\left(M-\bar{x}_w\right)}{M-m}\left[\left\|\Phi'\right\|_{\left[\bar{x}_w, M\right], \infty}+\left\|\Phi'\right\|_{\left[m, \bar{x}_w\right], \infty}\right] \\
&\leq \frac{1}{2}(M-m) \frac{\left\|\Phi'\right\|_{\left[\bar{w}_p, M\right], \infty}+\left\|\Phi'\right\|_{\left[m, \bar{w}_p\right], \infty}}{2} \leq \frac{1}{2}(M-m)\left\|\Phi'\right\|_{[m, M], \infty}
\end{aligned}$$

and

$$\begin{aligned}
(5.4)\qquad &\Theta_{\Phi}\left(\bar{x}_w; m, M\right) \\
&\leq \frac{1}{M-m}\left[\left(\bar{x}_w-m\right)\left(M-\bar{x}_w\right)^{1/q}\left\|\Phi'\right\|_{\left[\bar{x}_w, M\right], p}\right. \\
&\left.+\left(M-\bar{x}_w\right)\left(\bar{x}_w-m\right)^{1/q}\left\|\Phi'\right\|_{\left[m, \bar{x}_w\right], p}\right] \\
&\leq \frac{1}{M-m}\left[\left(\bar{x}_w-m\right)^q\left(M-\bar{x}_w\right)+\left(M-\bar{x}_w\right)^q\left(\bar{x}_w-m\right)\right]^{1/q}\left\|\Phi'\right\|_{[m, M], p}
\end{aligned}$$

where $p>1$, $\frac{1}{p}+\frac{1}{q}=1$.

PROOF. By the convexity of Φ we have that

$$\begin{aligned}
(5.5)\qquad &\sum_{i=1}^n w_i \Phi\left(x_i\right)-\Phi\left(\bar{x}_w\right) \\
&=\sum_{i=1}^n w_i \Phi\left[\frac{m\left(M-x_i\right)+M\left(x_i-m\right)}{M-m}\right]-\Phi\left(\bar{x}_w\right) \\
&\leq \sum_{i=1}^n w_i \frac{\left(M-x_i\right) \Phi(m)+\left(x_i-m\right) \Phi(M)}{M-m}-\Phi\left(\bar{x}_w\right) \\
&=\frac{\left(M-\bar{x}_w\right) \Phi(m)+\left(\bar{x}_w-m\right) \Phi(M)}{M-m}-\Phi\left(\bar{x}_w\right)=B .
\end{aligned}$$

By denoting

$$\Lambda_{\Phi}(t; m, M):=\frac{(t-m) \Phi(M)+(M-t) \Phi(m)}{M-m}-\Phi(t), \quad t \in[m, M]$$

we have

$$\begin{aligned}
(5.6)\qquad \Lambda_{\Phi}(t; m, M) &=\frac{(t-m) \Phi(M)+(M-t) \Phi(m)}{M-m}-\Phi(t) \\
&=\frac{(t-m) \Phi(M)+(M-t) \Phi(m)-(M-m) \Phi(t)}{M-m} \\
&=\frac{(t-m) \Phi(M)+(M-t) \Phi(m)-(M-t+t-m) \Phi(t)}{M-m} \\
&=\frac{(t-m)[\Phi(M)-\Phi(t)]-(M-t)[\Phi(t)-\Phi(m)]}{M-m}
\end{aligned}$$

for any $t \in[m, M]$. Also

$$B=\Lambda_{\Phi}\left(\bar{x}_w; m, M\right).$$

Taking the modulus on (5.6) and, noticing that, by the convexity of Φ we have

$$\begin{aligned}
&\Lambda_{\Phi}(t; m, M) \\
&=\frac{(t-m) \Phi(M)+(M-t) \Phi(m)}{M-m}-\Phi\left(\frac{(t-m) M+(M-t) m}{M-m}\right) \geq 0
\end{aligned}$$

for any $t \in [m, M]$, then we have

$$\Lambda_{\Phi}(t; m, M) \leq \frac{(t-m)|\Phi(M) - \Phi(t)| + (M-t)|\Phi(t) - \Phi(m)|}{M-m} \tag{5.7}$$
$$= \frac{(t-m)\left|\int_t^M \Phi'(s)\,ds\right| + (M-t)\left|\int_m^t \Phi'(s)\,ds\right|}{M-m}$$
$$\leq \frac{(t-m)\int_t^M |\Phi'(s)|\,ds + (M-t)\int_m^t |\Phi'(s)|\,ds}{M-m}$$

for any $t \in [m, M]$.

Finally, if we write the inequality (5.7) for $t = \bar{x}_w \in [m, M]$ and utilize the inequality (5.5), we deduce the desired result (5.1).

Now, we observe that

$$\frac{(t-m)\int_t^M |\Phi'(s)|\,ds + (M-t)\int_m^t |\Phi'(s)|\,ds}{M-m} \tag{5.8}$$
$$\leq \begin{cases} \max\{t-m, M-t\}\int_m^M |\Phi'(t)|\,dt \\ \\ \max\left\{\int_t^M |\Phi'(s)|\,ds, \int_m^t |\Phi'(s)|\,ds\right\}(M-m) \end{cases}$$
$$= \begin{cases} \left[\frac{1}{2}(M-m) + \left|t - \frac{m+M}{2}\right|\right]\int_m^M |\Phi'(t)|\,dt \\ \\ \left[\frac{1}{2}\int_m^M |\Phi'(s)|\,ds + \frac{1}{2}\left|\int_t^M |\Phi'(s)|\,ds - \int_m^t |\Phi'(s)|\,ds\right|\right](M-m) \end{cases}$$

for any $t \in [m, M]$. This proves the inequality (5.2).

By the Hölder's inequality we have

$$\int_t^M |\Phi'(s)|\,ds \leq \begin{cases} (M-t)\|\Phi'\|_{[t,M],\infty} \\ \\ (M-t)^{1/q}\|\Phi'\|_{[t,M],p} \quad \text{if } p>1, \frac{1}{p}+\frac{1}{q}=1 \end{cases}$$

and

$$\int_m^t |\Phi'(s)|\,ds \leq \begin{cases} (t-m)\|\Phi'\|_{[m,t],\infty} \\ \\ (t-m)^{1/q}\|\Phi'\|_{[m,t],p} \quad \text{if } p>1, \frac{1}{p}+\frac{1}{q}=1, \end{cases}$$

which give that

$$\frac{(t-m)\int_t^M |\Phi'(s)|\,ds + (M-t)\int_m^t |\Phi'(s)|\,ds}{M-m} \tag{5.9}$$
$$\leq \frac{(t-m)(M-t)\|\Phi'\|_{[t,M],\infty} + (M-t)(t-m)\|\Phi'\|_{[m,t],\infty}}{M-m}$$
$$= \frac{(t-m)(M-t)}{M-m}\left[\|\Phi'\|_{[t,M],\infty} + \|\Phi'\|_{[m,t],\infty}\right]$$
$$\leq \frac{1}{2}(M-m)\frac{\|\Phi'\|_{[t,M],\infty} + \|\Phi'\|_{[m,t],\infty}}{2}$$
$$\leq \frac{1}{2}(M-m)\max\left\{\|\Phi'\|_{[t,M],\infty}, \|\Phi'\|_{[m,t],\infty}\right\} = \frac{1}{2}(M-m)\|\Phi'\|_{[m,M],\infty}$$

and

$$\begin{aligned} (5.10) \qquad & \frac{(t-m)\int_t^M |\Phi'(s)|\,ds + (M-t)\int_m^t |\Phi'(s)|\,ds}{M-m} \\ & \leq \frac{(t-m)(M-t)^{1/q}\|\Phi'\|_{[t,M],p} + (M-t)(t-m)^{1/q}\|\Phi'\|_{[m,t],p}}{M-m} \\ & \leq \frac{1}{M-m}\left[\left((t-m)(M-t)^{1/q}\right)^q + \left((M-t)(t-m)^{1/q}\right)^q\right]^{1/q} \\ & \times \left[\|\Phi'\|_{[t,M],p}^p + \|\Phi'\|_{[m,t],p}^p\right]^{1/p} \\ & = \frac{1}{M-m}\left[(t-m)^q(M-t) + (M-t)^q(t-m)\right]^{1/q}\|\Phi'\|_{[m,M],p} \end{aligned}$$

for any $t \in [m, M]$.

These prove the desired inequalities (5.3) and (5.4). ∎

REMARK 5.1. Define the weighted arithmetic mean of the positive n-tuple $x = (x_1, ..., x_n)$ with the nonnegative weights $w = (w_1, ..., w_n)$ by

$$A_n(w, x) := \frac{1}{W_n}\sum_{i=1}^n w_i x_i$$

where $W_n := \sum_{i=1}^n w_i > 0$ and the weighted geometric mean of the same n-tuple, by

$$G_n(w, x) := \left(\prod_{i=1}^n x_i^{w_i}\right)^{1/W_n}.$$

It is well know that the following arithmetic mean-geometric mean inequality holds true

$$A_n(w, x) \geq G_n(w, x).$$

On applying the inequality (5.1) for the convex function $\Phi(t) = -\ln t$, we have the following reverse of the arithmetic mean-geometric mean inequality

$$(5.11) \qquad 1 \leq \frac{A_n(w,x)}{G_n(w,x)} \leq \left(\frac{A_n(w,x)}{m}\right)^{M-A_n(w,x)}\left(\frac{M}{A_n(w,x)}\right)^{A_n(w,x)-m}.$$

6. APPLICATIONS FOR THE HÖLDER INEQUALITY

If $x_i, y_i \geq 0$ for $i \in \{1, ..., n\}$, then the *Hölder inequality* holds true

$$\sum_{i=1}^n x_i y_i \leq \left(\sum_{i=1}^n x_i^p\right)^{1/p}\left(\sum_{i=1}^n y_i^q\right)^{1/q},$$

where $p > 1, \frac{1}{p} + \frac{1}{q} = 1$.

Assume that $p > 1$. If $z_i \in \mathbb{R}$ for $i \in \{1, ..., n\}$, satisfies the bounds

$$0 < m \leq z_i \leq M < \infty \text{ for } i \in \{1, ..., n\}$$

and $w_i \geq 0 \ \ (i = 1, \ldots, n)$ with $W_n := \sum_{i=1}^n w_i > 0$, then from Theorem 5.1 we have amongst other the following inequality

$$0 \leq \frac{\sum_{i=1}^n w_i z_i^p}{W_n} - \left(\frac{\sum_{i=1}^n w_i z_i}{W_n}\right)^p \leq (M^p - m^p)\left[\frac{1}{2} + \frac{1}{M-m}\left|\frac{\sum_{i=1}^n w_i z_i}{W_n} - \frac{m+M}{2}\right|\right]. \tag{6.1}$$

From this inequality we can state that:

PROPOSITION 6.1 (Dragomir, 2013 [**60**]). *If* $x_i \geq 0, y_i > 0$ *for* $i \in \{1, ..., n\}$, $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$ *and there exists the constants* $\gamma, \Gamma > 0$ *and such that*

$$\gamma \leq \frac{x_i}{y_i^{q-1}} \leq \Gamma \text{ for } i \in \{1, ..., n\},$$

then we have

$$0 \leq \frac{\sum_{i=1}^n x_i^p}{\sum_{i=1}^n y_i^q} - \left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q}\right)^p \leq (\Gamma^p - \gamma^p)\left[\frac{1}{2} + \frac{1}{\Gamma - \gamma}\left|\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q} - \frac{\gamma+\Gamma}{2}\right|\right]. \tag{6.2}$$

Finally, the following additive reverse of the Hölder inequality can be stated as well:

COROLLARY 6.2 (Dragomir, 2013 [**60**]). *With the assumptions of Proposition 6.1 we have*

$$0 \leq \left(\sum_{i=1}^n x_i^p\right)^{1/p}\left(\sum_{i=1}^n y_i^q\right)^{1/q} - \sum_{i=1}^n x_i y_i \leq (\Gamma^p - \gamma^p)^{1/p}\left[\frac{1}{2} + \frac{1}{\Gamma - \gamma}\left|\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^q} - \frac{\gamma+\Gamma}{2}\right|\right]^{1/p} \sum_{i=1}^n y_i^q. \tag{6.3}$$

REMARK 6.1. We observe that for $p = q = 2$ we have from the first inequality in (6.2) the following reverse of the Cauchy-Bunyakovsky-Schwarz inequality

$$\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n x_i y_i\right)^2 \leq \left(\Gamma^2 - \gamma^2\right)\left[\frac{1}{2} + \frac{1}{\Gamma - \gamma}\left|\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n y_i^2} - \frac{\gamma+\Gamma}{2}\right|\right]\left(\sum_{i=1}^n y_i^2\right)^2 \tag{6.4}$$

provided that $x_i \geq 0$, $y_i > 0$ for $i \in \{1, ..., n\}$ and there exists the constants $\gamma, \Gamma > 0$ such that

$$\gamma \leq \frac{x_i}{y_i} \leq \Gamma \; for \; i \in \{1, ..., n\}.$$

7. APPLICATIONS FOR f-DIVERGENCE

Consider the f-divergence

$$I_f(p, q) := \sum_{i=1}^n q_i f\left(\frac{p_i}{q_i}\right) \tag{7.1}$$

defined on the set of probability distributions $p, q \in \mathbb{P}^n$, where f is convex on $(0, \infty)$. It is assumed that $f(u)$ is zero and strictly convex at $u = 1$. By appropriately defining this convex function, various divergences are derived.

PROPOSITION 7.1 (Dragomir, 2013 [**60**]). *Let $f:(0,\infty)\to\mathbb{R}$ be a convex function with the property that $f(1)=0$. Assume that $p,q\in\mathbb{P}^n$ and there exists the constants $0<r<1<R<\infty$ such that*

$$r\leq\frac{p_i}{q_i}\leq R \text{ for } i\in\{1,...,n\}. \tag{7.2}$$

Then we have the inequalities

$$0\leq I_f(p,q)\leq B_f(r,R) \tag{7.3}$$

where

$$B_f(r,R):=\frac{(R-1)\int_r^1|f'(t)|\,dt+(1-r)\int_1^R|f'(t)|\,dt}{R-r}. \tag{7.4}$$

Moreover, we have the following bounds for $B_f(r,R)$

$$\begin{aligned} &B_f(r,R)\\ &\leq\begin{cases}\left[\frac{1}{2}+\frac{\left|1-\frac{r+R}{2}\right|}{R-r}\right]\int_r^R|f'(t)|\,dt\\ \\ \left[\frac{1}{2}\int_r^R|f'(t)|\,dt+\frac{1}{2}\left|\int_1^R|f'(t)|\,dt-\int_r^1|f'(t)|\,dt\right|\right],\end{cases}\end{aligned} \tag{7.5}$$

and

$$\begin{aligned} &B_f(r,R)\\ &\leq\frac{(1-r)(R-1)}{R-r}\left[\|f'\|_{[1,R],\infty}+\|f'\|_{[r,1],\infty}\right]\\ &\leq\frac{1}{2}(R-r)\frac{\|f'\|_{[1,R],\infty}+\|f'\|_{[r,1],\infty}}{2}\leq\frac{1}{2}(R-r)\|f'\|_{[r,R],\infty}\end{aligned} \tag{7.6}$$

and

$$\begin{aligned} &B_f(r,R)\\ &\leq\frac{1}{R-r}\left[(1-r)(R-1)^{1/q}\|f'\|_{[1,R],p}+(R-1)(1-r)^{1/q}\|f'\|_{[r,1],p}\right]\\ &\leq\frac{1}{R-r}\left[(1-r)^q(R-1)+(R-1)^q(1-r)\right]^{1/q}\|f'\|_{[r,R],p}\end{aligned} \tag{7.7}$$

where $p>1$, $\frac{1}{p}+\frac{1}{q}=1$.

The proof follows by Theorem 5.1 by choosing $w_i=q_i$, $x_i=\frac{p_i}{q_i}$, $m=r$ and $M=R$ and performing the required calculations. The details are omitted.

The above results can be utilized to obtain various inequalities for the divergence measures in information theory that are particular instances of f-divergence.

Consider the Kullback-Leibler divergence

$$KL(p,q)=\sum_{i=1}^n p_i\log\left(\frac{p_i}{q_i}\right),\ p,q\in\mathbb{P}^n.$$

For the convex function $f:(0,\infty)\to\mathbb{R}$, $f(t)=-\ln t$ we have

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

If $p, q \in \mathbb{P}^n$ such that there exists the constants $0 < r < 1 < R < \infty$ with

$$r \leq \frac{p_i}{q_i} \leq R \text{ for } i \in \{1, ..., n\}, \tag{7.8}$$

then we get from the inequality (7.4)

$$0 \leq KL(q, p) \leq \ln \left(\frac{R^{1-r}}{r^{R-1}} \right)^{\frac{1}{R-r}}. \tag{7.9}$$

For $\alpha > 1$, let

$$f(t) = t^\alpha, \ t > 0.$$

Then

$$I_f(p, q) = D_\alpha(p, q) = \sum_{i=1}^n p_i^\alpha q_i^{1-\alpha},$$

which is the α-order entropy.

If $p, q \in \mathbb{P}^n$ such that (7.8) holds true, then by (7.4) we have

$$0 \leq D_\alpha(p, q) \leq \frac{(R-1)(1-r^\alpha) + (1-r)(R^\alpha - 1)}{R-r}.$$

8. A Refinement and a New Reverse

For a real function $g : [m, M] \to \mathbb{R}$ and two distinct points $\alpha, \beta \in [m, M]$ we recall that the *divided difference* of g in these points is defined by

$$[\alpha, \beta; g] := \frac{g(\beta) - g(\alpha)}{\beta - \alpha}.$$

THEOREM 8.1 (Dragomir, 2011 [**56**]). *Let $f : I \to \mathbb{R}$ be a continuous convex function on the interval of real numbers I and $m, M \in \mathbb{R}, m < M$ with $[m, M] \subset \mathring{I}$, $\mathring{I}$ the interior of I. Let $\bar{\mathbf{a}} = (a_1, \ldots, a_n), \bar{\mathbf{p}} = (p_1, \ldots, p_n)$ be $n-$tuples of real numbers with $p_i \geq 0$ $(i \in \{1, \ldots, n\})$ and $\sum_{i=1}^n p_i = 1$. If $m \leq a_i \leq M, \ \ i \in \{1, \ldots, n\}$, with $\sum_{i=1}^n p_i a_i \neq m, M$, then*

$$\begin{aligned} &\left| \sum_{i=1}^n p_i \left| f(a_i) - f\left(\sum_{j=1}^n p_j a_j\right) \right| sgn\left(a_i - \sum_{j=1}^n p_j a_j\right) \right| \\ &\leq \sum_{i=1}^n p_i f(a_i) - f\left(\sum_{i=1}^n p_i a_i\right) \\ &\leq \frac{1}{2}\left(\left[\sum_{i=1}^n p_i a_i, M; f\right] - \left[m, \sum_{i=1}^n p_i a_i; f\right]\right) \sum_{i=1}^n p_i \left| a_i - \sum_{j=1}^n p_j a_j \right| \\ &\leq \frac{1}{2}\left(\left[\sum_{i=1}^n p_i a_i, M; f\right] - \left[m, \sum_{i=1}^n p_i a_i; f\right]\right) \left[\sum_{i=1}^n p_i a_i^2 - \left(\sum_{j=1}^n p_j a_j\right)^2\right]^{1/2}. \end{aligned} \tag{8.1}$$

If the lateral derivatives $f'_+(m)$ and $f'_-(M)$ are finite, then we also have the inequalities

$$0 \le \sum_{i=1}^{n} p_i f(a_i) - f\left(\sum_{i=1}^{n} p_i a_i\right) \tag{8.2}$$
$$\le \frac{1}{2}\left(\left[\sum_{i=1}^{n} p_i a_i, M; f\right] - \left[m, \sum_{i=1}^{n} p_i a_i; f\right]\right) \sum_{i=1}^{n} p_i \left|a_i - \sum_{j=1}^{n} p_j a_j\right|$$
$$\le \frac{1}{2}\left[f'_-(M) - f'_+(m)\right] \sum_{i=1}^{n} p_i \left|a_i - \sum_{j=1}^{n} p_j a_j\right|$$
$$\le \frac{1}{2}\left[f'_-(M) - f'_+(m)\right] \left[\sum_{i=1}^{n} p_i a_i^2 - \left(\sum_{j=1}^{n} p_j a_j\right)^2\right]^{1/2}.$$

PROOF. We recall that if $f : I \to \mathbb{R}$ is a continuous convex function on the interval of real numbers I and $\alpha \in I$ then the *divided difference function* $f_\alpha : I \setminus \{\alpha\} \to \mathbb{R}$,

$$f_\alpha(t) := [\alpha, t; f] := \frac{f(t) - f(\alpha)}{t - \alpha}$$

is monotonic nondecreasing on $I \setminus \{\alpha\}$.

For $\bar{a}_p := \sum_{j=1}^{n} p_j a_j \in (m, M)$, we consider now the sequence

$$f_{\bar{a}_p}(i) := \frac{f(a_i) - f(\bar{a}_p)}{a_i - \bar{a}_p}.$$

We will show that $f_{\bar{a}_p}(i)$ and $h_i := a_i - \bar{a}_p, \in \{1, \dots, n\}$ are synchronous.

Let $i, j \in \{1, \dots, n\}$ with $a_i, a_j \ne \bar{a}_p$. Assume that $a_i \ge a_j$, then by the monotonicity of f_α we have

$$f_{\bar{a}_p}(i) = \frac{f(a_i) - f(\bar{a}_p)}{a_i - \bar{a}_p} \ge \frac{f(a_j) - f(\bar{a}_p)}{a_j - \bar{a}_p} = f_{\bar{a}_p}(j) \tag{8.3}$$

and

$$h_i \ge h_j \tag{8.4}$$

which shows that

$$\left[f_{\bar{a}_p}(i) - f_{\bar{a}_p}(j)\right](h_i - h_j) \ge 0. \tag{8.5}$$

If $a_i < a_j$, then the inequalities (8.3) and (8.4) reverse but the inequality (8.5) still holds true.

Utilising the continuity property of the modulus we have

$$\left|\left[\left|f_{\bar{a}_p}(i)\right| - \left|f_{\bar{a}_p}(j)\right|\right](h_i - h_j)\right| \le \left|\left[f_{\bar{a}_p}(i) - f_{\bar{a}_p}(j)\right](h_i - h_j)\right|$$
$$= \left[f_{\bar{a}_p}(i) - f_{\bar{a}_p}(j)\right](h_i - h_j)$$

for any $i, j \in \{1, \dots, n\}$.

Multiplying with $p_i, p_j \geq 0$ and summing over i and j from 1 to n we have

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \left[\left|f_{\bar{a}_p}(i)\right| - \left|f_{\bar{a}_p}(j)\right|\right](h_i - h_j)\right| \leq \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \left[f_{\bar{a}_p}(i) - f_{\bar{a}_p}(j)\right](h_i - h_j). \tag{8.6}$$

A simple calculation shows that

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \left[\left|f_{\bar{a}_p}(i)\right| - \left|f_{\bar{a}_p}(j)\right|\right](h_i - h_j) \\
&= \sum_{i=1}^{n} p_i \left|f_{\bar{a}_p}(i)\right| h_i - \sum_{i=1}^{n} p_i \left|f_{\bar{a}_p}(i)\right| \sum_{i=1}^{n} p_i h_i \\
&= \sum_{i=1}^{n} p_i \left|\frac{f(a_i) - f(\bar{a}_p)}{a_i - \bar{a}_p}\right|(a_i - \bar{a}_p) \\
&- \sum_{i=1}^{n} p_i \left|\frac{f(a_i) - f(\bar{a}_p)}{a_i - \bar{a}_p}\right| \sum_{i=1}^{n} p_i (a_i - \bar{a}_p) \\
&= \sum_{i=1}^{n} p_i \left|\frac{f(a_i) - f(\bar{a}_p)}{a_i - \bar{a}_p}\right|(a_i - \bar{a}_p) \\
&= \sum_{i=1}^{n} p_i \left|f(a_i) - f(\bar{a}_p)\right| sgn(a_i - \bar{a}_p)
\end{aligned} \tag{8.7}$$

and

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \left[f_{\bar{a}_p}(i) - f_{\bar{a}_p}(j)\right](h_i - h_j) \\
&= \sum_{i=1}^{n} p_i f_{\bar{a}_p}(i) h_i - \sum_{i=1}^{n} p_i f_{\bar{a}_p}(i) \sum_{i=1}^{n} p_i h_i \\
&= \sum_{i=1}^{n} p_i \left(\frac{f(a_i) - f(\bar{a}_p)}{a_i - \bar{a}_p}\right)(a_i - \bar{a}_p) \\
&- \sum_{i=1}^{n} p_i \left(\frac{f(a_i) - f(\bar{a}_p)}{a_i - \bar{a}_p}\right) \sum_{i=1}^{n} p_i (a_i - \bar{a}_p) \\
&= \sum_{i=1}^{n} p_i \left(\frac{f(a_i) - f(\bar{a}_p)}{a_i - \bar{a}_p}\right)(a_i - \bar{a}_p) \\
&= \sum_{i=1}^{n} p_i f(a_i) - f\left(\sum_{i=1}^{n} p_i a_i\right).
\end{aligned} \tag{8.8}$$

On making use of the identities (8.7) and (8.8) we obtain from (8.6) the first inequality in (8.1).

Now, since $\bar{a}_p := \sum_{j=1}^{n} p_j a_j \in (m, M)$ then we have by the monotonicity of $f_{\bar{a}_p}(i)$ that

$$[m, \bar{a}_p; f] = \frac{f(\bar{a}_p) - f(m)}{\bar{a}_p - m} \leq f_{\bar{a}_p}(i) \leq \frac{f(M) - f(\bar{a}_p)}{M - \bar{a}_p} = [\bar{a}_p, M; f] \tag{8.9}$$

for any $i \in \{1, \ldots, n\}$.

Applying now the *Grüss' type inequality* obtained by Cerone & Dragomir in [**18**]

$$\left|\sum_{i=1}^{n} w_i x_i y_i - \sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i y_i\right| \leq \frac{1}{2}(\Gamma - \gamma) \sum_{i=1}^{n} w_i \left|x_i - \sum_{j=1}^{n} w_j x_j\right|$$

provided

$$-\infty < \delta \leq y_i \leq \Delta < \infty \tag{8.10}$$

for $i = 1, \ldots, n$, we have that

$$\sum_{i=1}^{n} p_i f(a_i) - f\left(\sum_{i=1}^{n} p_i a_i\right) \leq \frac{1}{2}([\bar{a}_p, M; f] - [m, \bar{a}_p; f]) \sum_{i=1}^{n} p_i \left|a_i - \sum_{j=1}^{n} p_j a_j\right|,$$

which proves the second inequality in (8.1).

The last bound in (8.1) is obvious by Cauchy-Bunyakovsky-Schwarz discrete inequality.

If the lateral derivatives $f'_+(m)$ and $f'_-(M)$ are finite, then by the convexity of f we have the *gradient inequalities*

$$\frac{f(M) - f(\bar{a}_p)}{M - \bar{a}_p} \leq f'_-(M)$$

and

$$\frac{f(\bar{a}_p) - f(m)}{\bar{a}_p - m} \geq f'_+(m),$$

where $\bar{a}_p \in (m, M)$. These imply that

$$[\bar{a}_p, M; f] - [m, \bar{a}_p; f] \leq f'_-(M) - f'_+(m)$$

and the proof of the third inequality in (8.2) is concluded.

The rest is obvious. ∎

REMARK 8.1. Define the weighted arithmetic mean of the positive n-tuple $x = (x_1, ..., x_n)$ with the nonnegative weights $w = (w_1, ..., w_n)$ by

$$A_n(w, x) := \frac{1}{W_n} \sum_{i=1}^{n} w_i x_i$$

where $W_n := \sum_{i=1}^{n} w_i > 0$ and the weighted geometric mean of the same n-tuple, by

$$G_n(w, x) := \left(\prod_{i=1}^{n} x_i^{w_i}\right)^{1/W_n}.$$

It is well know that the following arithmetic mean-geometric mean inequality holds

$$A_n(w, x) \geq G_n(w, x).$$

Applying the inequality (8.2) for the convex function $f(t) = -\ln t, t > 0$ we have the following reverse of the arithmetic mean-geometric mean inequality

$$
\begin{aligned}
(8.11) \qquad 1 &\le \frac{A_n(w,x)}{G_n(w,x)} \\
&\le \left[\frac{\left(\frac{A_n(w,x)}{m}\right)^{A_n(w,x)-m}}{\left(\frac{M}{A_n(w,x)}\right)^{M-A_n(w,x)}} \right]^{\frac{1}{2}A_n(w,|x-A_n(w,x)|)} \\
&\le \exp\left[\frac{1}{2}\frac{M-m}{mM} A_n(w,|x-A_n(w,x)|)\right],
\end{aligned}
$$

provided that $0 < m \le x_i \le M < \infty$ for $i \in \{1, ..., n\}$.

9. APPLICATIONS FOR THE HÖLDER INEQUALITY

If $x_i, y_i \ge 0$ for $i \in \{1, ..., n\}$, then the *Hölder inequality* holds true

$$\sum_{i=1}^{n} x_i y_i \le \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1/q},$$

where $p > 1, \frac{1}{p} + \frac{1}{q} = 1$.

Assume that $p > 1$. If $z_i \in \mathbb{R}$ for $i \in \{1, ..., n\}$, satisfies the bounds

$$0 < m \le z_i \le M < \infty \text{ for } i \in \{1, ..., n\}$$

and $w_i \ge 0 \ \ (i = 1, \dots, n)$ with $W_n := \sum_{i=1}^{n} w_i > 0$, then from Theorem 8.1 we have amongst other the following inequality

$$
\begin{aligned}
(9.1) \qquad & \left| \frac{1}{W_n} \sum_{i=1}^{n} \left| z_i^p - \left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n} \right)^p \right| w_i sgn\left[z_i - \frac{\sum_{i=1}^{n} w_i z_i}{W_n} \right] d\mu \right| \\
&\le \frac{\sum_{i=1}^{n} w_i z_i^p}{W_n} - \left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n} \right)^p \\
&\le \frac{1}{2} \left(\left[\frac{\sum_{i=1}^{n} w_i z_i}{W_n}, M; (\cdot)^p \right] - \left[m, \frac{\sum_{i=1}^{n} w_i z_i}{W_n}; (\cdot)^p \right] \right) \tilde{D}_w(z) \\
&\le \frac{1}{2} \left(\left[\frac{\sum_{i=1}^{n} w_i z_i}{W_n}, M; (\cdot)^p \right] - \left[m, \frac{\sum_{i=1}^{n} w_i z_i}{W_n}; (\cdot)^p \right] \right) \tilde{D}_{w,2}(z) \\
&\le \frac{1}{4} \left(\left[\frac{\sum_{i=1}^{n} w_i z_i}{W_n}, M; (\cdot)^p \right] - \left[m, \frac{\sum_{i=1}^{n} w_i z_i}{W_n}; (\cdot)^p \right] \right) (M - m),
\end{aligned}
$$

where $\frac{\sum_{i=1}^{n} w_i z_i}{W_n} \in (m, M)$ and

$$\tilde{D}_w(z) := \frac{1}{W_n} \sum_{i=1}^{n} w_i \left| z_i - \frac{\sum_{j=1}^{n} w_j z_j}{W_n} \right|$$

while

$$\tilde{D}_{w,2}(z) = \left[\frac{\sum_{i=1}^{n} w_i z_i^2}{W_n} - \left(\frac{\sum_{i=1}^{n} w_i z_i}{W_n} \right)^2 \right]^{\frac{1}{2}}.$$

The following result related to the Hölder inequality holds:

PROPOSITION 9.1 (Dragomir, 2011 [**56**]). *If* $x_i \geq 0, y_i > 0$ *for* $i \in \{1,...,n\}$, $p > 1$, $\frac{1}{p}+\frac{1}{q} = 1$ *and there exists the constants* $\gamma, \Gamma > 0$ *and such that*

$$\gamma \leq \frac{x_i}{y_i^{q-1}} \leq \Gamma \text{ for } i \in \{1,...,n\},$$

then we have

$$\left|\sum_{i=1}^{n}\left|\frac{x_i^p}{y_i^q} - \left(\frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} y_j^q}\right)^p\right| y_i^q sgn\left[\frac{x_i}{y_i^{q-1}} - \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} y_j^q}\right]\right| \tag{9.2}$$
$$\leq \frac{\sum_{i=1}^{n} x_i^p}{\sum_{i=1}^{n} y_i^q} - \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}\right)^p$$
$$\leq \frac{1}{2}\left(\left[\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}, \Gamma; (\cdot)^p\right] - \left[\gamma, \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}; (\cdot)^p\right]\right) \tilde{D}_{y^q}\left(\frac{x}{y^{q-1}}\right)$$
$$\leq \frac{1}{2}\left(\left[\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}, \Gamma; (\cdot)^p\right] - \left[\gamma, \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}; (\cdot)^p\right]\right) \tilde{D}_{y^q,2}\left(\frac{x}{y^{q-1}}\right)$$
$$\leq \frac{1}{4}\left(\left[\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}, \Gamma; (\cdot)^p\right] - \left[\gamma, \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^q}; (\cdot)^p\right]\right) (\Gamma - \gamma),$$

where

$$\tilde{D}_{y^q}\left(\frac{x}{y^{q-1}}\right) = \frac{1}{\sum_{i=1}^{n} y_i^q}\sum_{i=1}^{n} y_i^q\left|\frac{x_i}{y_i^{q-1}} - \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} y_j^q}\right|$$

and

$$\tilde{D}_{y^q,2}\left(\frac{x}{y^{q-1}}\right) = \left[\frac{1}{\sum_{i=1}^{n} y_i^q}\sum_{i=1}^{n}\frac{x_i^2}{y_i^{q-2}} - \left(\frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} y_j^q}\right)^2\right]^{\frac{1}{2}}.$$

PROOF. The inequalities (9.3) follow from (9.1) by choosing

$$z_i = \frac{x_i}{y_i^{q-1}} \text{ and } w_i = y_j^q.$$

The details are omitted. ∎

REMARK 9.1. We observe that for $p = q = 2$ we have from the first inequality in (9.2) the following reverse of the Cauchy-Bunyakovsky-Schwarz inequality

$$\left|\sum_{i=1}^{n}\left|\frac{x_i^2}{y_i^2} - \left(\frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} y_j^2}\right)^2\right| y_i^2 sgn\left(\frac{x_i}{y_i} - \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} y_j^2}\right)\right| \tag{9.3}$$
$$\leq \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} y_i^2} - \left(\frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^2}\right)^2$$
$$\leq \frac{1}{2}(\Gamma - \gamma)\frac{1}{\sum_{i=1}^{n} y_i^2}\sum_{i=1}^{n} y_i^2\left|\frac{x_i}{y_i} - \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} y_j^2}\right|$$
$$\leq \frac{1}{2}(\Gamma - \gamma)\left[\frac{1}{\sum_{i=1}^{n} y_i^2}\sum_{i=1}^{n} x_i^2 - \left(\frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} y_j^2}\right)^2\right]^{\frac{1}{2}}$$
$$\leq \frac{1}{4}(\Gamma - \gamma)^2,$$

provided that there exists the constants $\gamma, \Gamma > 0$ such that

$$\gamma \leq \frac{x_i}{y_i} \leq \Gamma \text{ for } i \in \{1, ..., n\}.$$

10. APPLICATIONS FOR f-DIVERGENCE

Consider the f-*divergence*

$$I_f(p,q) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \tag{10.1}$$

defined on the set of probability distributions $p, q \in \mathbb{P}^n$, where f is convex on $(0, \infty)$. It is assumed that $f(u)$ is zero and strictly convex at $u = 1$.

PROPOSITION 10.1 (Dragomir, 2011 [**56**]). *Let $f : (0, \infty) \to \mathbb{R}$ be a convex function with the property that $f(1) = 0$. Assume that $p, q \in \mathbb{P}^n$ and there exists the constants $0 < r < 1 < R < \infty$ such that*

$$r \leq \frac{p_i}{q_i} \leq R \text{ for } i \in \{1, ..., n\}. \tag{10.2}$$

Then we have

$$\begin{aligned} 0 \leq I_f(p,q) &\leq \frac{1}{2}([1, R; f] - [r, 1; f]) D_v(p,q) \\ &\leq \frac{1}{2}\left[f'_-(R) - f'_+(r)\right] D_v(p,q) \\ &\leq \frac{1}{2}\left[f'_-(R) - f'_+(r)\right] \left[D_{\chi^2}(p,q)\right]^{1/2} \\ &\leq \frac{1}{4}(R - r)\left[f'_-(R) - f'_+(r)\right], \end{aligned} \tag{10.3}$$

where $D_v(p,q) = \sum_{i=1}^{n} |p_i - q_i|$ and $D_{\chi^2}(p,q) = \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1$.

PROOF. From (8.2) we have

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) - f(1) \\ &\leq \frac{1}{2}([1, R; f] - [r, 1; f]) \sum_{i=1}^{n} q_i \left|\frac{p_i}{q_i} - 1\right| \\ &\leq \frac{1}{2}\left[f'_-(R) - f'_+(r)\right] \sum_{i=1}^{n} q_i \left|\frac{p_i}{q_i} - 1\right| \\ &\leq \frac{1}{2}\left[f'_-(R) - f'_+(r)\right] \left(\sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1\right)^{1/2} \leq \frac{1}{4}(R - r)\left[f'_-(R) - f'_+(r)\right] \end{aligned}$$

i.e., the desired result (10.3). ∎

REMARK 10.1. The inequality

$$I_f(p,q) \leq \frac{1}{4}(R - r)\left[f'_-(R) - f'_+(r)\right] \tag{10.4}$$

was obtained for the discrete divergence measures in 2000 by S.S. Dragomir, see [**46**].

PROPOSITION 10.2 (Dragomir, 2011 [**56**]). *With the assumptions in Proposition 10.1 we have*

$$\left|I_{|f|(sgn(\cdot)-1)}(p,q)\right| \le I_f(p,q) \le \frac{1}{2}\left([1,R;f]-[r,1;f]\right)D_v(p,q) \le \frac{1}{2}\left([1,R;f]-[r,1;f]\right)\left[D_{\chi^2}(p,q)\right]^{1/2} \le \frac{1}{4}\left([1,R;f]-[r,1;f]\right)(R-r), \tag{10.5}$$

where $I_{|f|(sgn(\cdot)-1)}(p,q)$ *is the generalized* f*-divergence for the non-necessarily convex function* $|f|(sgn(\cdot)-1)$ *and is defined by*

$$I_{|f|(sgn(\cdot)-1)}(p,q) := \sum_{i=1}^{n} q_i \left|f\left(\frac{p_i}{q_i}\right)\right| sgn\left(\frac{p_i}{q_i}-1\right). \tag{10.6}$$

PROOF. From the inequality (8.1) we have

$$\begin{aligned} &\left|\sum_{i=1}^{n} q_i \left|f\left(\frac{p_i}{q_i}\right)-f(1)\right| sgn\left(\frac{p_i}{q_i}-1\right)\right| \\ &\le \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right)-f(1) \\ &\le \frac{1}{2}\left([1,R;f]-[r,1;f]\right)\sum_{i=1}^{n} q_i\left|\frac{p_i}{q_i}-1\right| \\ &\le \frac{1}{2}\left([1,R;f]-[r,1;f]\right)\left(\sum_{i=1}^{n}\frac{p_i^2}{q_i}-1\right)^{1/2} \\ &\le \frac{1}{4}\left([1,R;f]-[r,1;f]\right)(R-r) \end{aligned}$$

from where we get the desired result (10.5). ∎

The above results can be utilized to obtain various inequalities for the divergence measures in information theory that are particular instances of f-divergence.

Consider the *Kullback-Leibler divergence*

$$KL(p,q) = \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{q_i}\right),\ p,q \in \mathbb{P}^n.$$

For the convex function $f:(0,\infty)\to\mathbb{R}$, $f(t) = -\ln t$ we have

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

If $p,q\in\mathbb{P}^n$ such that there exists the constants $0<r<1<R<\infty$ with

$$r \le \frac{p_i}{q_i} \le R \text{ for } i \in \{1,...,n\}, \tag{10.7}$$

then we get from the first inequality in (10.3) that

$$0 \le KL(q,p) \le \frac{1}{2} D_v(p,q) \ln\left(\frac{1}{R^{R-1}r^{1-r}}\right).$$

For the convex function $f : (0, \infty) \to \mathbb{R}$, $f(t) = t \ln t$ we have

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

If $p, q \in \mathbb{P}^n$ are such that there exists the constants $0 < r < 1 < R < \infty$ with the property (10.7), then we get from the first inequality in (10.3) that

$$0 \le KL(p, q) \le \frac{1}{2} D_v(p, q) \ln\left(R^{\frac{R}{R-1}} r^{\frac{r}{1-r}}\right).$$

CHAPTER 9

Refinements of Jensen's Inequality

1. Introduction

The Jensen inequality for convex functions plays a crucial role in the Theory of Inequalities due to the fact that other inequalities such as that arithmetic mean-geometric mean inequality, Hölder and Minkowski inequalities, Ky Fan's inequality etc. can be obtained as particular cases of it.

Let C be a convex subset of the linear space X and f a convex function on C. If $\mathbf{p} = (p_1, \dots, p_n)$ is a probability sequence and $\mathbf{x} = (x_1, \dots, x_n) \in C^n$, then

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i), \tag{1.1}$$

is well known in the literature as Jensen's inequality.

In 1989, J. Pečarić and the author [**112**] obtained the following refinement of (1.1):

$$\begin{aligned} f\left(\sum_{i=1}^{n} p_i x_i\right) &\leq \sum_{i_1,\dots,i_{k+1}=1}^{n} p_{i_1} \dots p_{i_{k+1}} f\left(\frac{x_{i_1} + \cdots + x_{i_{k+1}}}{k+1}\right) \\ &\leq \sum_{i_1,\dots,i_k=1}^{n} p_{i_1} \dots p_{i_k} f\left(\frac{x_{i_1} + \cdots + x_{i_k}}{k}\right) \\ &\leq \cdots \leq \sum_{i=1}^{n} p_i f(x_i), \end{aligned} \tag{1.2}$$

for $k \geq 1$ and $\mathbf{p}$,$\mathbf{x}$ as above.

If $q_1, \dots, q_k \geq 0$ with $\sum_{j=1}^{k} q_j = 1$, then the following refinement obtained in 1994 by the author [**28**] also holds:

$$\begin{aligned} f\left(\sum_{i=1}^{n} p_i x_i\right) &\leq \sum_{i_1,\dots,i_k=1}^{n} p_{i_1} \dots p_{i_k} f\left(\frac{x_{i_1} + \cdots + x_{i_k}}{k}\right) \\ &\leq \sum_{i_1,\dots,i_k=1}^{n} p_{i_1} \dots p_{i_k} f\left(q_1 x_{i_1} + \cdots + q_k x_{i_k}\right) \\ &\leq \sum_{i=1}^{n} p_i f(x_i), \end{aligned} \tag{1.3}$$

where $1 \leq k \leq n$ and $\mathbf{p}$, $\mathbf{x}$ are as above.

For other refinements and applications related to Ky Fan's inequality, the arithmetic mean-geometric mean inequality, the generalised triangle inequality etc., see [**29**]-[**75**].

2. General Results

The following result may be stated.

THEOREM 2.1 (Dragomir, 2010 [**54**]). *Let* $f : C \to \mathbb{R}$ *be a convex function on the convex subset* C *of the linear space* X, $x_i \in C$, $p_i > 0$, $i \in \{1, \dots, n\}$ *with* $\sum_{i=1}^{n} p_i = 1$. *Then*

$$f\left(\sum_{j=1}^{n} p_j x_j\right) \le \min_{k\in\{1,\dots,n\}} \left[(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + p_k f(x_k)\right] \tag{2.1}$$
$$\le \frac{1}{n}\left[\sum_{k=1}^{n} (1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + \sum_{k=1}^{n} p_k f(x_k)\right]$$
$$\le \max_{k\in\{1,\dots,n\}} \left[(1-p_k) f\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right) + p_k f(x_k)\right]$$
$$\le \sum_{j=1}^{n} p_j f(x_j).$$

In particular,

$$f\left(\frac{1}{n}\sum_{j=1}^{n} x_j\right) \le \frac{1}{n}\min_{k\in\{1,\dots,n\}} \left[(n-1) f\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right) + f(x_k)\right] \tag{2.2}$$
$$\le \frac{1}{n^2}\left[(n-1)\sum_{k=1}^{n} f\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right) + \sum_{k=1}^{n} f(x_k)\right]$$

$$\le \frac{1}{n}\max_{k\in\{1,\dots,n\}} \left[(n-1) f\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right) + f(x_k)\right]$$
$$\le \frac{1}{n}\sum_{j=1}^{n} f(x_j).$$

PROOF. For any $k \in \{1, \dots, n\}$, we have

$$\sum_{j=1}^{n} p_j x_j - p_k x_k = \sum_{\substack{j=1\\ j\neq k}}^{n} p_j x_j = \frac{\sum_{\substack{j=1\\ j\neq k}}^{n} p_j}{\sum_{\substack{j=1\\ j\neq k}}^{n} p_j} \sum_{\substack{j=1\\ j\neq k}}^{n} p_j x_j = (1-p_k)\cdot \frac{1}{\sum_{\substack{j=1\\ j\neq k}}^{n} p_j} \sum_{\substack{j=1\\ j\neq k}}^{n} p_j x_j$$

which implies that

$$\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k} = \frac{1}{\sum_{\substack{j=1\\ j\neq k}}^{n} p_j} \sum_{\substack{j=1\\ j\neq k}}^{n} p_j x_j \in C \tag{2.3}$$

for each $k \in \{1, \dots, n\}$, since the right side of (2.3) is a convex combination of the elements $x_j \in C$, $j \in \{1, \dots, n\} \setminus \{k\}$.

Taking the function f on (2.3) and applying the Jensen inequality, we get successively

$$f\left(\frac{\sum_{j=1}^{n}p_jx_j-p_kx_k}{1-p_k}\right)=f\left(\frac{1}{\sum\limits_{\substack{j=1\\j\neq k}}^{n}p_j}\sum_{\substack{j=1\\j\neq k}}^{n}p_jx_j\right)\le\frac{1}{\sum\limits_{\substack{j=1\\j\neq k}}^{n}p_j}\sum_{\substack{j=1\\j\neq k}}^{n}p_jf\left(x_j\right)$$
$$=\frac{1}{1-p_k}\left[\sum_{j=1}^{n}p_jf\left(x_j\right)-p_kf\left(x_k\right)\right]$$

for any $k\in\{1,\dots,n\}$, which implies

$$(1-p_k)\,f\left(\frac{\sum_{j=1}^{n}p_jx_j-p_kx_k}{1-p_k}\right)+p_kf\left(x_k\right)\le\sum_{j=1}^{n}p_jf\left(x_j\right) \tag{2.4}$$

for each $k\in\{1,\dots,n\}$.

Utilising the convexity of f, we also have

$$\begin{aligned}(1-p_k)\,f&\left(\frac{\sum_{j=1}^{n}p_jx_j-p_kx_k}{1-p_k}\right)+p_kf\left(x_k\right)\\&\ge f\left[(1-p_k)\cdot\frac{\sum_{j=1}^{n}p_jx_j-p_kx_k}{1-p_k}+p_kx_k\right]=f\left(\sum_{j=1}^{n}p_jx_j\right)\end{aligned} \tag{2.5}$$

for each $k\in\{1,\dots,n\}$.

Taking the minimum over k in (2.5), utilising the fact that

$$\min_{k\in\{1,\dots,n\}}\alpha_k\le\frac{1}{n}\sum_{k=1}^{n}\alpha_k\le\max_{k\in\{1,\dots,n\}}\alpha_k$$

and then taking the maximum in (2.4), we deduce the desired inequality (2.1). ∎

After setting $x_j=y_j-\sum_{l=1}^{n}q_ly_l$ and $p_j=q_j, j\in\{1,\dots,n\}$, Theorem 2.1 becomes the following corollary:

COROLLARY 2.2 (Dragomir, 2010 [**54**]). *Let $f:C\to\mathbb{R}$ be a convex function on the convex subset C, $0\in C$, $y_j\in X$ and $q_j>0$, $j\in\{1,\dots,n\}$ with $\sum_{j=1}^{n}q_j=1$. If $y_j-\sum_{l=1}^{n}q_ly_l\in C$ for any $j\in\{1,\dots,n\}$, then*

$$\begin{aligned}&f\left(0\right)\\&\le\min_{k\in\{1,\dots,n\}}\left\{(1-q_k)\,f\left[\frac{q_k}{1-q_k}\left(\sum_{l=1}^{n}q_ly_l-y_k\right)\right]+q_kf\left(y_k-\sum_{l=1}^{n}q_ly_l\right)\right\}\\&\le\frac{1}{n}\left\{\sum_{l=1}^{n}(1-q_k)\,f\left[\frac{q_k}{1-q_k}\left(\sum_{l=1}^{n}q_ly_l-y_k\right)\right]+\sum_{l=1}^{n}q_kf\left(y_k-\sum_{l=1}^{n}q_ly_l\right)\right\}\\&\le\max_{k\in\{1,\dots,n\}}\left\{(1-q_k)\,f\left[\frac{q_k}{1-q_k}\left(\sum_{l=1}^{n}q_ly_l-y_k\right)\right]+q_kf\left(y_k-\sum_{l=1}^{n}q_ly_l\right)\right\}\\&\le\sum_{j=1}^{n}q_jf\left(y_j-\sum_{l=1}^{n}q_ly_l\right).\end{aligned} \tag{2.6}$$

In particular, if $y_j - \frac{1}{n}\sum_{l=1}^{n} y_l \in C$ *for any* $j \in \{1,\dots,n\}$, *then*

$$f(0) \leq \frac{1}{n}\min_{k\in\{1,\dots,n\}}\left\{(n-1)f\left[\frac{1}{n-1}\left(\frac{1}{n}\sum_{l=1}^{n}y_l - y_k\right)\right] + f\left(y_k - \frac{1}{n}\sum_{l=1}^{n}y_l\right)\right\}$$
$$\leq \frac{1}{n^2}\left\{(n-1)\sum_{k=1}^{n}f\left[\frac{1}{n-1}\left(\frac{1}{n}\sum_{l=1}^{n}y_l - y_k\right)\right] + \sum_{k=1}^{n}f\left(y_k - \frac{1}{n}\sum_{l=1}^{n}y_l\right)\right\}$$
$$\leq \frac{1}{n}\max_{k\in\{1,\dots,n\}}\left\{(n-1)f\left[\frac{1}{n-1}\left(\frac{1}{n}\sum_{l=1}^{n}y_l - y_k\right)\right] + f\left(y_k - \frac{1}{n}\sum_{l=1}^{n}y_l\right)\right\}$$
$$\leq \frac{1}{n}\sum_{j=1}^{n}f\left(y_j - \frac{1}{n}\sum_{l=1}^{n}y_l\right). \tag{2.7}$$

The above results can be applied for various convex functions related to celebrated inequalities as mentioned in the introduction.

Application 1. If $(X, \|\cdot\|)$ is a normed linear space and $p \geq 1$, then the function $f : X \to \mathbb{R}$, $f(x) = \|x\|^p$ is convex on X. Now, on applying Theorem 2.1 and Corollary 2.2 for $x_i \in X$, $p_i > 0$, $i \in \{1,\dots,n\}$ with $\sum_{i=1}^{n} p_i = 1$, we get:

$$\left\|\sum_{j=1}^{n}p_jx_j\right\|^p \leq \min_{k\in\{1,\dots,n\}}\left[(1-p_k)^{1-p}\left\|\sum_{j=1}^{n}p_jx_j - p_kx_k\right\|^p + p_k\|x_k\|^p\right]$$
$$\leq \frac{1}{n}\left[\sum_{k=1}^{n}(1-p_k)^{1-p}\left\|\sum_{j=1}^{n}p_jx_j - p_kx_k\right\|^p + \sum_{k=1}^{n}p_k\|x_k\|^p\right]$$
$$\leq \max_{k\in\{1,\dots,n\}}\left[(1-p_k)^{1-p}\left\|\sum_{j=1}^{n}p_jx_j - p_kx_k\right\|^p + p_k\|x_k\|^p\right]$$
$$\leq \sum_{j=1}^{n}p_j\|x_j\|^p \tag{2.8}$$

and

$$\max_{k\in\{1,\dots,n\}}\left\{\left[(1-p_k)^{1-p}p_k^p + p_k\right]\left\|x_k - \sum_{l=1}^{n}p_lx_l\right\|^p\right\} \leq \sum_{j=1}^{n}p_j\left\|x_k - \sum_{l=1}^{n}p_lx_l\right\|^p. \tag{2.9}$$

In particular, we have the inequality:

$$\left\|\frac{1}{n}\sum_{j=1}^{n}x_j\right\|^p \le \frac{1}{n}\min_{k\in\{1,\dots,n\}}\left[(n-1)^{1-p}\left\|\sum_{j=1}^{n}x_j - x_k\right\|^p + \|x_k\|^p\right] \tag{2.10}$$
$$\le \frac{1}{n^2}\left[(n-1)^{1-p}\sum_{k=1}^{n}\left\|\sum_{j=1}^{n}x_j - x_k\right\|^p + \sum_{k=1}^{n}\|x_k\|^p\right]$$
$$\le \frac{1}{n}\max_{k\in\{1,\dots,n\}}\left[(n-1)^{1-p}\left\|\sum_{j=1}^{n}x_j - x_k\right\|^p + \|x_k\|^p\right]$$
$$\le \frac{1}{n}\sum_{j=1}^{n}\|x_j\|^p$$

and

$$\left[(n-1)^{1-p}+1\right]\max_{k\in\{1,\dots,n\}}\left\|x_k - \frac{1}{n}\sum_{l=1}^{n}x_l\right\|^p \le \sum_{j=1}^{n}\left\|x_j - \frac{1}{n}\sum_{l=1}^{n}x_l\right\|^p. \tag{2.11}$$

If we consider the function $h_p(t) := (1-t)^{1-p}t^p + t$, $p \ge 1$, $t \in [0,1)$, then we observe that

$$h_p'(t) = 1 + pt^{p-1}(1-t)^{1-p} + (p-1)t^p(1-t)^{-p},$$

which shows that h_p is strictly increasing on $[0,1)$. Therefore,

$$\min_{k\in\{1,\dots,n\}}\left\{(1-p_k)^{1-p}p_k^p + p_k\right\} = p_m + (1-p_m)^{1-p}p_m^p,$$

where $p_m := \min_{k\in\{1,\dots,n\}} p_k$. By (2.9), we then obtain the following inequality:

$$\left[p_m + (1-p_m)^{1-p}\cdot p_m^p\right]\max_{k\in\{1,\dots,n\}}\left\|x_k - \sum_{l=1}^{n}p_l x_l\right\|^p \tag{2.12}$$
$$\le \sum_{j=1}^{n}p_j\left\|x_j - \sum_{l=1}^{n}p_l x_l\right\|^p.$$

Application 2. Let $x_i, p_i > 0$, $i \in \{1,\dots,n\}$ with $\sum_{i=1}^{n}p_i = 1$. The following inequality is well known in the literature as the *arithmetic mean-geometric mean* inequality:

$$\sum_{j=1}^{n}p_j x_j \ge \prod_{j=1}^{n}x_j^{p_j}. \tag{2.13}$$

The equality case holds in (2.13) iff $x_1 = \cdots = x_n$.

Applying the inequality (2.1) for the convex function $f:(0,\infty)\to\mathbb{R}$, $f(x)=-\ln x$ and performing the necessary computations, we derive the following refinement of (2.13):

$$\begin{aligned}(2.14)\qquad \sum_{i=1}^{n} p_i x_i &\geq \max_{k\in\{1,\ldots,n\}}\left\{\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right)^{1-p_k}\cdot x_k^{p_k}\right\}\\ &\geq \prod_{k=1}^{n}\left[\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right)^{1-p_k}\cdot x_k^{p_k}\right]^{\frac{1}{n}}\\ &\geq \min_{k\in\{1,\ldots,n\}}\left\{\left(\frac{\sum_{j=1}^{n} p_j x_j - p_k x_k}{1-p_k}\right)^{1-p_k}\cdot x_k^{p_k}\right\}\geq \prod_{i=1}^{n} x_i^{p_i}.\end{aligned}$$

In particular, we have the inequality:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} x_i &\geq \max_{k\in\{1,\ldots,n\}}\left\{\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right)^{\frac{n-1}{n}}\cdot x_k^{\frac{1}{n}}\right\}\\ &\geq \prod_{k=1}^{n}\left[\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right)^{\frac{n-1}{n}}\cdot x_k^{\frac{1}{n}}\right]^{\frac{1}{n}}\\ &\geq \min_{k\in\{1,\ldots,n\}}\left\{\left(\frac{\sum_{j=1}^{n} x_j - x_k}{n-1}\right)^{\frac{n-1}{n}}\cdot x_k^{\frac{1}{n}}\right\}\geq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}.\end{aligned}$$

3. APPLICATIONS FOR f-DIVERGENCES

The following refinement of the positivity property of f-divergence may be stated.

THEOREM 3.1 (Dragomir, 2010 [**54**]). *For any* $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$, *we have the inequalities*

$$\begin{aligned}(3.1)\qquad I_f(\mathbf{p},\mathbf{q}) &\geq \max_{k\in\{1,\ldots,n\}}\left[(1-q_k) f\left(\frac{1-p_k}{1-q_k}\right)+q_k f\left(\frac{p_k}{q_k}\right)\right]\\ &\geq \frac{1}{n}\left[\sum_{k=1}^{n}(1-q_k) f\left(\frac{1-p_k}{1-q_k}\right)+\sum_{k=1}^{n} q_k f\left(\frac{p_k}{q_k}\right)\right]\\ &\geq \min_{k\in\{1,\ldots,n\}}\left[(1-q_k) f\left(\frac{1-p_k}{1-q_k}\right)+q_k f\left(\frac{p_k}{q_k}\right)\right]\geq 0,\end{aligned}$$

provided $f:[0,\infty)\to\mathbb{R}$ *is convex and normalized on* $[0,\infty)$.

The proof is obvious by Theorem 2.1 applied for the convex function $f:[0,\infty)\to\mathbb{R}$ and for the choice $x_i=\frac{p_i}{q_i}$, $i\in\{1,\ldots,n\}$ and the probabilities q_i, $i\in\{1,\ldots,n\}$.

If we consider a new divergence measure $R_f(\mathbf{p},\mathbf{q})$ defined for $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ by

$$(3.2)\qquad R_f(\mathbf{p},\mathbf{q}) := \frac{1}{n-1}\sum_{k=1}^{n}(1-q_k) f\left(\frac{1-p_k}{1-q_k}\right)$$

and call it the *reverse* $f-$*divergence*, we observe that

$$(3.3)\qquad R_f(\mathbf{p},\mathbf{q}) = I_f(\mathbf{r},\mathbf{t})$$

with

$$\mathbf{r} = \left(\frac{1-p_1}{n-1}, \dots, \frac{1-p_n}{n-1}\right), \quad \mathbf{t} = \left(\frac{1-q_1}{n-1}, \dots, \frac{1-q_n}{n-1}\right) \qquad (n \geq 2).$$

With this notation, we can state the following corollary of the above proposition.

COROLLARY 3.2. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, *we have*

$$I_f(\mathbf{p}, \mathbf{q}) \geq R_f(\mathbf{p}, \mathbf{q}) \geq 0. \tag{3.4}$$

The proof is obvious by the second inequality in (3.1) and the details are omitted.

In what follows, we point out some particular inequalities for various instances of divergence measures such as: the *total variation distance*, χ^2*-divergence*, *Kullback-Leibler divergence*, *Jeffreys divergence*.

The *total variation distance* is defined by the convex function $f(t) = |t-1|$, $t \in \mathbb{R}$ and given in:

$$V(p,q) := \sum_{j=1}^{n} q_j \left|\frac{p_j}{q_j} - 1\right| = \sum_{j=1}^{n} |p_j - q_j|. \tag{3.5}$$

The following improvement of the positivity inequality for the total variation distance can be stated as follows.

PROPOSITION 3.3. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, *we have the inequality:*

$$V(p,q) \geq 2 \max_{k \in \{1,\dots,n\}} |p_k - q_k| \quad (\geq 0). \tag{3.6}$$

The proof follows by the first inequality in (3.1) for $f(t) = |t-1|$, $t \in \mathbb{R}$.

The K. Pearson χ^2*-divergence* is obtained for the convex function $f(t) = (1-t)^2$, $t \in \mathbb{R}$ and given by

$$\chi^2(p,q) := \sum_{j=1}^{n} q_j \left(\frac{p_j}{q_j} - 1\right)^2 = \sum_{j=1}^{n} \frac{(p_j - q_j)^2}{q_j}. \tag{3.7}$$

PROPOSITION 3.4. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$,

$$\chi^2(p,q) \geq \max_{k \in \{1,\dots,n\}} \left\{\frac{(p_k - q_k)^2}{q_k(1-q_k)}\right\} \geq 4 \max_{k \in \{1,\dots,n\}} (p_k - q_k)^2 \quad (\geq 0). \tag{3.8}$$

PROOF. On applying the first inequality in (3.1) for the function $f(t) = (1-t)^2$, $t \in \mathbb{R}$, we get

$$\begin{aligned} \chi^2(p,q) &\geq \max_{k \in \{1,\dots,n\}} \left\{(1-q_k)\left(\frac{1-p_k}{1-q_k} - 1\right)^2 + q_k\left(\frac{p_k}{q_k} - 1\right)^2\right\} \\ &= \max_{k \in \{1,\dots,n\}} \left\{\frac{(p_k - q_k)^2}{q_k(1-q_k)}\right\}. \end{aligned}$$

Since

$$q_k(1-q_k) \leq \frac{1}{4}[q_k + (1-q_k)]^2 = \frac{1}{4},$$

then

$$\frac{(p_k - q_k)^2}{q_k(1-q_k)} \geq 4(p_k - q_k)^2$$

for each $k \in \{1, \dots, n\}$, which proves the last part of (3.8). ∎

The *Kullback-Leibler divergence* can be obtained for the convex function $f : (0, \infty) \to \mathbb{R}$, $f(t) = t \ln t$ and is defined by

$$KL(p, q) := \sum_{j=1}^{n} q_j \cdot \frac{p_j}{q_j} \ln \left(\frac{p_j}{q_j} \right) = \sum_{j=1}^{n} p_j \ln \left(\frac{p_j}{q_j} \right). \tag{3.9}$$

PROPOSITION 3.5. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, *we have:*

$$KL(p, q) \geq \ln \left[\max_{k \in \{1,\dots,n\}} \left\{ \left(\frac{1 - p_k}{1 - q_k} \right)^{1 - p_k} \cdot \left(\frac{p_k}{q_k} \right)^{p_k} \right\} \right] \geq 0. \tag{3.10}$$

PROOF. The first inequality is obvious by Theorem 3.1. Utilising the inequality between the *geometric mean and the harmonic mean,*

$$x^{\alpha} y^{1-\alpha} \geq \frac{1}{\frac{\alpha}{x} + \frac{1-\alpha}{y}}, \qquad x, y > 0, \ \alpha \in [0, 1]$$

we have

$$\left(\frac{1 - p_k}{1 - q_k} \right)^{1 - p_k} \cdot \left(\frac{p_k}{q_k} \right)^{p_k} \geq 1,$$

for any $k \in \{1, \dots, n\}$, which implies the second part of (3.10). ∎

Another divergence measure that is of importance in Information Theory is the *Jeffreys divergence*

$$J(p, q) := \sum_{j=1}^{n} q_j \cdot \left(\frac{p_j}{q_j} - 1 \right) \ln \left(\frac{p_j}{q_j} \right) = \sum_{j=1}^{n} (p_j - q_j) \ln \left(\frac{p_j}{q_j} \right), \tag{3.11}$$

which is an f-divergence for $f(t) = (t - 1) \ln t$, $t > 0$.

PROPOSITION 3.6. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, *we have:*

$$\begin{aligned} J(p, q) &\geq \max_{k \in \{1,\dots,n\}} \left\{ (q_k - p_k) \ln \left[\frac{(1 - p_k) q_k}{(1 - q_k) p_k} \right] \right\} \\ &\geq \max_{k \in \{1,\dots,n\}} \left[\frac{(q_k - p_k)^2}{p_k + q_k - 2 p_k q_k} \right] \geq 0. \end{aligned} \tag{3.12}$$

PROOF. Writing the first inequality in Theorem 3.1 for $f(t) = (t - 1) \ln t$, we have

$$\begin{aligned} J(p, q) &\geq \max_{k \in \{1,\dots,n\}} \left\{ (1 - q_k) \left[\left(\frac{1 - p_k}{1 - q_k} - 1 \right) \ln \left(\frac{1 - p_k}{1 - q_k} \right) \right] + q_k \left(\frac{p_k}{q_k} - 1 \right) \ln \left(\frac{p_k}{q_k} \right) \right\} \\ &= \max_{k \in \{1,\dots,n\}} \left\{ (q_k - p_k) \ln \left(\frac{1 - p_k}{1 - q_k} \right) - (q_k - p_k) \ln \left(\frac{p_k}{q_k} \right) \right\} \\ &= \max_{k \in \{1,\dots,n\}} \left\{ (q_k - p_k) \ln \left[\frac{(1 - p_k) q_k}{(1 - q_k) p_k} \right] \right\}, \end{aligned}$$

proving the first inequality in (3.12).

Utilising the elementary inequality for positive numbers,

$$\frac{\ln b - \ln a}{b - a} \geq \frac{2}{a + b}, \qquad a, b > 0$$

we have

$$\begin{aligned}
&(q_k - p_k)\left[\ln\left(\frac{1-p_k}{1-q_k}\right) - \ln\left(\frac{p_k}{q_k}\right)\right] \\
&= (q_k - p_k)\cdot\frac{\ln\left(\frac{1-p_k}{1-q_k}\right) - \ln\left(\frac{p_k}{q_k}\right)}{\frac{1-p_k}{1-q_k} - \frac{p_k}{q_k}}\cdot\left[\frac{1-p_k}{1-q_k} - \frac{p_k}{q_k}\right] \\
&= \frac{(q_k - p_k)^2}{q_k(1-q_k)}\cdot\frac{\ln\left(\frac{1-p_k}{1-q_k}\right) - \ln\left(\frac{p_k}{q_k}\right)}{\frac{1-p_k}{1-q_k} - \frac{p_k}{q_k}} \\
&\geq \frac{(q_k - p_k)^2}{q_k(1-q_k)}\cdot\frac{2}{\frac{1-p_k}{1-q_k} + \frac{p_k}{q_k}} = \frac{2(q_k - p_k)^2}{p_k + q_k - 2p_kq_k} \geq 0,
\end{aligned}$$

for each $k \in \{1, \ldots, n\}$, giving the second inequality in (3.12). ∎

4. MORE GENERAL RESULTS

Let C be a convex subset in the real linear space X and assume that $f : C \to \mathbb{R}$ is a convex function on C. If $x_i \in C$ and $p_i > 0, i \in \{1, ..., n\}$ with $\sum_{i=1}^n p_i = 1$, then for any nonempty subset J of $\{1, ..., n\}$ we put $\bar{J} := \{1, ..., n\} \setminus J\, (\neq \emptyset)$ and define $P_J := \sum_{i\in J} p_i$ and $\bar{P}_J := P_{\bar{J}} = \sum_{j\in\bar{J}} p_j = 1 - \sum_{i\in J} p_i$. For the convex function f and the n-tuples $\mathbf{x} := (x_1, ..., x_n)$ and $\mathbf{p} := (p_1, ..., p_n)$ as above, we can define the following functional

$$D(f, \mathbf{p}, \mathbf{x}; J) := P_J f\left(\frac{1}{P_J}\sum_{i\in J} p_i x_i\right) + \bar{P}_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}} p_j x_j\right) \tag{4.1}$$

where here and everywhere below $J \subset \{1, ..., n\}$ with $J \neq \emptyset$ and $J \neq \{1, ..., n\}$.

It is worth to observe that for $J = \{k\}, k \in \{1, ..., n\}$ we have the functional

$$\begin{aligned}
D_k(f, \mathbf{p}, \mathbf{x}) &:= D(f, \mathbf{p}, \mathbf{x}; \{k\}) \\
&= p_k f(x_k) + (1 - p_k) f\left(\frac{\sum_{i=1}^n p_i x_i - p_k x_k}{1 - p_k}\right)
\end{aligned} \tag{4.2}$$

that has been investigated in the earlier paper [**54**].

THEOREM 4.1 (Dragomir, 2010 [**55**]). *Let C be a convex subset in the real linear space X and assume that $f : C \to \mathbb{R}$ is a convex function on C. If $x_i \in C$ and $p_i > 0, i \in \{1, ..., n\}$ with $\sum_{i=1}^n p_i = 1$ then for any nonempty subset J of $\{1, ..., n\}$ we have*

$$\sum_{k=1}^n p_k f(x_k) \geq D(f, \mathbf{p}, \mathbf{x}; J) \geq f\left(\sum_{k=1}^n p_k x_k\right). \tag{4.3}$$

PROOF. By the convexity of the function f we have

$$\begin{aligned}D(f,\mathbf{p},\mathbf{x};J) &= P_J f\left(\frac{1}{P_J}\sum_{i\in J} p_i x_i\right) + \bar{P}_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j x_j\right) \\ &\geq f\left[P_J\left(\frac{1}{P_J}\sum_{i\in J} p_i x_i\right) + \bar{P}_J\left(\frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j x_j\right)\right] \\ &= f\left(\sum_{k=1}^{n} p_k x_k\right)\end{aligned}$$

for any J, which proves the second inequality in (4.3).

By the Jensen inequality we also have

$$\begin{aligned}\sum_{k=1}^{n} p_k f(x_k) &= \sum_{i\in J} p_i f(x_i) + \sum_{j\in \bar{J}} p_j f(x_j) \\ &\geq P_J f\left(\frac{1}{P_J}\sum_{i\in J} p_i x_i\right) + \bar{P}_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j x_j\right) \\ &= D(f,\mathbf{p},\mathbf{x};J)\end{aligned}$$

for any J, which proves the first inequality in (4.3). ∎

REMARK 4.1. We observe that the inequality (4.3) can be written in an equivalent form as

$$\sum_{k=1}^{n} p_k f(x_k) \geq \max_{\emptyset\neq J\subset\{1,\dots,n\}} D(f,\mathbf{p},\mathbf{x};J) \tag{4.4}$$

and

$$\min_{\emptyset\neq J\subset\{1,\dots,n\}} D(f,\mathbf{p},\mathbf{x};J) \geq f\left(\sum_{k=1}^{n} p_k x_k\right). \tag{4.5}$$

These inequalities imply the following results that have been obtained earlier by the author in [**54**] utilising a different method of proof slightly more complicated:

$$\sum_{k=1}^{n} p_k f(x_k) \geq \max_{k\in\{1,\dots,n\}} D_k(f,\mathbf{p},\mathbf{x}) \tag{4.6}$$

and

$$\min_{k\in\{1,\dots,n\}} D_k(f,\mathbf{p},\mathbf{x}) \geq f\left(\sum_{k=1}^{n} p_k x_k\right). \tag{4.7}$$

Moreover, since

$$\max_{\emptyset\neq J\subset\{1,\dots,n\}} D(f,\mathbf{p},\mathbf{x};J) \geq \max_{k\in\{1,\dots,n\}} D_k(f,\mathbf{p},\mathbf{x})$$

and

$$\min_{k\in\{1,\dots,n\}} D_k(f,\mathbf{p},\mathbf{x}) \geq \min_{\emptyset\neq J\subset\{1,\dots,n\}} D(f,\mathbf{p},\mathbf{x};J),$$

then the new inequalities (4.4) and (4.4) are better than the earlier results developed in [**54**].

The case of uniform distribution, namely, when $p_i = \frac{1}{n}$ for all $\{1, ..., n\}$ is of interest as well. If we consider a natural number m with $1 \le m \le n-1$ and if we define

$$D_m(f, \mathbf{x}) := \frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right) \tag{4.8}$$

then we can state the following result:

COROLLARY 4.2 (Dragomir, 2010 [**55**]). *Let C be a convex subset in the real linear space X and assume that $f : C \to \mathbb{R}$ is a convex function on C. If $x_i \in C$, then for any $m \in \{1, ..., n-1\}$ we have*

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) \ge D_m(f, \mathbf{x}) \ge f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right). \tag{4.9}$$

In particular, we have the bounds

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k) \ge \max_{m \in \{1,...,n-1\}} \left[\frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right)\right] \tag{4.10}$$

and

$$\min_{m \in \{1,...,n-1\}} \left[\frac{m}{n} f\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right) + \frac{n-m}{n} f\left(\frac{1}{n-m}\sum_{j=m+1}^{n} x_j\right)\right] \ge f\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right). \tag{4.11}$$

The following version of the inequality (4.3) may be useful for symmetric convex functions:

COROLLARY 4.3 (Dragomir, 2010 [**55**]). *Let C be a convex function with the property that $0 \in C$. If $y_j \in X$ such that for $p_i > 0, i \in \{1, ..., n\}$ with $\sum_{i=1}^{n} p_i = 1$ we have $y_j - \sum_{i=1}^{n} p_i y_i \in C$ for any $j \in \{1, ..., n\}$, then for any nonempty subset J of $\{1, ..., n\}$ we have*

$$\sum_{k=1}^{n} p_k f\left(y_k - \sum_{i=1}^{n} p_i y_i\right) \ge P_J f\left[\bar{P}_J\left(\frac{1}{P_J}\sum_{i\in J} p_i y_i - \frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j y_j\right)\right] + \bar{P}_J f\left[P_J\left(\frac{1}{\bar{P}_J}\sum_{j\in \bar{J}} p_j y_j - \frac{1}{P_J}\sum_{i\in J} p_i y_i\right)\right] \ge f(0). \tag{4.12}$$

REMARK 4.2. If C is as in Corollary 4.3 and $y_j \in X$ such that $y_j - \frac{1}{n}\sum_{i=1}^{n} y_i \in C$ for any $j \in \{1, ..., n\}$ then for any $m \in \{1, ..., n-1\}$ we have

$$\frac{1}{n}\sum_{k=1}^{n} f\left(y_k - \frac{1}{n}\sum_{i=1}^{n} y_i\right) \ge \frac{m}{n} f\left[\frac{n-m}{n}\left(\frac{1}{m}\sum_{i=1}^{m} y_i - \frac{1}{n-m}\sum_{j=m+1}^{n} y_j\right)\right] + \frac{n-m}{n} f\left[\frac{m}{n}\left(\frac{1}{n-m}\sum_{j=m+1}^{n} y_j - \frac{1}{m}\sum_{i=1}^{m} y_i\right)\right] \ge f(0). \tag{4.13}$$

REMARK 4.3. It is also useful to remark that if $J = \{k\}$ where $k \in \{1, ..., n\}$ then the particular form we can derive from (4.12) can be written as

$$\sum_{\ell=1}^{n} p_\ell f\left(y_\ell - \sum_{i=1}^{n} p_i y_i\right) \geq p_k f\left[(1-p_k)\left(y_k - \frac{1}{1-p_k}\left(\sum_{j=1}^{n} p_j y_j - p_k y_k\right)\right)\right] + (1-p_k) f\left[p_k\left(\frac{1}{1-p_k}\left(\sum_{j=1}^{n} p_j y_j - p_k y_k\right) - y_k\right)\right] \geq f(0) \tag{4.14}$$

which is equivalent with

$$\sum_{\ell=1}^{n} p_\ell f\left(y_\ell - \sum_{i=1}^{n} p_i y_i\right) \geq p_k f\left(y_k - \sum_{j=1}^{n} p_j y_j\right) + (1-p_k) f\left[\frac{p_k}{1-p_k}\left(\sum_{j=1}^{n} p_j y_j - y_k\right)\right] \geq f(0) \tag{4.15}$$

for any $k \in \{1, ..., n\}$.

5. A LOWER BOUND FOR MEAN f-DEVIATION

Let X be a real linear space. For a convex function $f : X \to \mathbb{R}$ with the properties that $f(0) = 0$, define the *mean f-deviation* of an n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ with the probability distribution $\mathbf{p} = (p_1, ..., p_n)$ by the non-negative quantity

$$K_f(\mathbf{p}, \mathbf{x}) := \sum_{i=1}^{n} p_i f\left(x_i - \sum_{k=1}^{n} p_k x_k\right). \tag{5.1}$$

The fact that $K_f(\mathbf{p}, \mathbf{x})$ is non-negative follows by Jensen's inequality, namely

$$K_f(\mathbf{p}, \mathbf{x}) \geq f\left(\sum_{i=1}^{n} p_i \left(x_i - \sum_{k=1}^{n} p_k x_k\right)\right) = f(0) = 0.$$

A natural example of such deviations can be provided by the convex function $f(x) := \|x\|^r$ with $r \geq 1$ defined on a normed linear space $(X, \|\cdot\|)$. We denote this by

$$K_r(\mathbf{p}, \mathbf{x}) := \sum_{i=1}^{n} p_i \left\|x_i - \sum_{k=1}^{n} p_k x_k\right\|^r \tag{5.2}$$

and call it the *mean r-absolute deviation* of the n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ with the probability distribution $\mathbf{p} = (p_1, ..., p_n)$.

The following result that provides a lower bound for the mean f-deviation holds:

THEOREM 5.1 (Dragomir, 2010 [**55**]). *Let $f : X \to [0, \infty)$ be a convex function with $f(0) = 0$. If $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and $\mathbf{p} = (p_1, ..., p_n)$ is a probability distribution with all p_i*

nonzero, then

$$K_f(\mathbf{p},\mathbf{x}) \geq \max_{\emptyset\neq J\subset\{1,\dots,n\}} \left\{ P_J f\left[\bar{P}_J\left(\frac{1}{P_J}\sum_{i\in J}p_ix_i - \frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_jx_j\right)\right] + P_J f\left(\frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_jy_j - \frac{1}{P_J}\sum_{i\in J}p_iy_i\right)\right\} (\geq 0). \tag{5.3}$$

In particular, we have

$$K_f(\mathbf{p},\mathbf{x}) \geq \max_{k\in\{1,\dots,n\}} \left\{ (1-p_k) f\left[\frac{p_k}{1-p_k}\left(\sum_{l=1}^n p_lx_l - x_k\right)\right] + p_k f\left(x_k - \sum_{l=1}^n p_lx_l\right)\right\} (\geq 0). \tag{5.4}$$

The proof follows from Corollary 4.3 and Remark 4.3.

As a particular case of interest, we have the following:

COROLLARY 5.2 (Dragomir, 2010 [**55**]). *Let* $(X,\|\cdot\|)$ *be a normed linear space. If* $\mathbf{x}=(x_1,\dots,x_n)\in X^n$ *and* $\mathbf{p}=(p_1,\dots,p_n)$ *is a probability distribution with all* p_i *nonzero, then for* $r\geq 1$ *we have*

$$K_r(\mathbf{p},\mathbf{x}) \geq \max_{\emptyset\neq J\subset\{1,\dots,n\}} \left\{ P_J\bar{P}_J\left(\bar{P}_J^{r-1}+P_J^{r-1}\right)\left\|\frac{1}{P_J}\sum_{i\in J}p_ix_i - \frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_jx_j\right\|^r\right\} (\geq 0). \tag{5.5}$$

REMARK 5.1. By the convexity of the power function $f(t)=t^r, r\geq 1$ we have

$$\begin{aligned} P_J\bar{P}_J\left(\bar{P}_J^{r-1}+P_J^{r-1}\right) &= P_J\bar{P}_J^r + \bar{P}_JP_J^r \\ &\geq \left(P_J\bar{P}_J+\bar{P}_JP_J\right)^r = 2^rP_J^r\bar{P}_J^r \end{aligned}$$

therefore

$$\begin{aligned} &P_J\bar{P}_J\left(\bar{P}_J^{r-1}+P_J^{r-1}\right)\left\|\frac{1}{P_J}\sum_{i\in J}p_ix_i - \frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_jx_j\right\|^r \\ &\geq 2^rP_J^r\bar{P}_J^r\left\|\frac{1}{P_J}\sum_{i\in J}p_ix_i - \frac{1}{\bar{P}_J}\sum_{j\in\bar{J}}p_jx_j\right\|^r = 2^r\left\|\bar{P}_J\sum_{i\in J}p_ix_i - P_J\sum_{j\in\bar{J}}p_jx_j\right\|^r. \end{aligned} \tag{5.6}$$

Since

$$\begin{aligned} \bar{P}_J\sum_{i\in J}p_ix_i - P_J\sum_{j\in\bar{J}}p_jx_j &= (1-P_J)\sum_{i\in J}p_ix_i - P_J\left(\sum_{k=1}^n p_kx_k - \sum_{i\in J}p_ix_i\right) \\ &= \sum_{i\in J}p_ix_i - P_J\sum_{k=1}^n p_kx_k, \end{aligned} \tag{5.7}$$

then by (5.5)-(5.7) we deduce the coarser but perhaps more useful lower bound

$$K_r(\mathbf{p},\mathbf{x}) \geq 2^r \max_{\emptyset\neq J\subset\{1,\dots,n\}} \left\{\left\|\sum_{i\in J}p_ix_i - P_J\sum_{k=1}^n p_kx_k\right\|^r\right\} (\geq 0). \tag{5.8}$$

The case for mean r-absolute deviation is incorporated in:

COROLLARY 5.3 (Dragomir, 2010 [**55**]). *Let* $(X, \|\cdot\|)$ *be a normed linear space. If* $\mathbf{x} = (x_1, ..., x_n) \in X^n$ *and* $\mathbf{p} = (p_1, ..., p_n)$ *is a probability distribution with all* p_i *nonzero, then for* $r \geq 1$ *we have*

$$K_r(\mathbf{p}, \mathbf{x}) \geq \max_{k \in \{1,...,n\}} \left\{ \left[(1 - p_k)^{1-r} p_k^r + p_k \right] \left\| x_k - \sum_{l=1}^{n} p_l x_l \right\|^r \right\}. \tag{5.9}$$

REMARK 5.2. Since the function $h_r(t) := (1-t)^{1-r} t^r + t$, $r \geq 1$, $t \in [0, 1)$ is strictly increasing on $[0, 1)$, then

$$\min_{k \in \{1,...,n\}} \left\{ (1 - p_k)^{1-r} p_k^r + p_k \right\} = p_m + (1 - p_m)^{1-r} p_m^r,$$

where $p_m := \min_{k \in \{1,...,n\}} p_k$.

We then obtain the following simpler inequality:

$$K_r(\mathbf{p}, \mathbf{x}) \geq \left[p_m + (1 - p_m)^{1-r} \cdot p_m^r \right] \max_{k \in \{1,...,n\}} \left\| x_k - \sum_{l=1}^{n} p_l x_l \right\|^p, \tag{5.10}$$

which is perhaps more useful for applications(see also [**54**]).

6. APPLICATIONS FOR f-DIVERGENCE MEASURES

We endeavour to extend the concept of f-divergence for functions defined on a cone in a linear space as follows.

Firstly, we recall that the subset K in a linear space X is a *cone* if the following two conditions are satisfied:

(i) for any $x, y \in K$ we have $x + y \in K$;

(ii) for any $x \in K$ and any $\alpha \geq 0$ we have $\alpha x \in K$.

For a given n-tuple of vectors $\mathbf{z} = (z_1, ..., z_n) \in K^n$ and a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero, we can define, for the convex function $f : K \to \mathbb{R}$, the following *f-divergence of* $\mathbf{z}$ *with the distribution* $\mathbf{q}$

$$I_f(\mathbf{z}, \mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{z_i}{q_i}\right). \tag{6.1}$$

It is obvious that if $X = \mathbb{R}$, $K = [0, \infty)$ and $\mathbf{x} = \mathbf{p} \in \mathbb{P}^n$ then we obtain the usual concept of the f-divergence associated with a function $f : [0, \infty) \to \mathbb{R}$.

Now, for a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1, ..., n\}$ we have

$$\mathbf{q}_J := \left(Q_J, \bar{Q}_J\right) \in \mathbb{P}^2$$

and

$$\mathbf{x}_J := \left(X_J, \bar{X}_J\right) \in K^2$$

where, as above

$$X_J := \sum_{i \in J} x_i, \quad \text{and} \quad \bar{X}_J := X_{\bar{J}}.$$

It is obvious that

$$I_f(\mathbf{x}_J, \mathbf{q}_J) = Q_J f\left(\frac{X_J}{Q_J}\right) + \bar{Q}_J f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right).$$

The following inequality for the f-divergence of an n-tuple of vectors in a linear space holds:

THEOREM 6.1 (Dragomir, 2010 [**55**]). *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1, ..., n\}$ we have*

$$\begin{aligned} I_f(\mathbf{x}, \mathbf{q}) &\geq \max_{\emptyset \neq J \subset \{1,...,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq I_f(\mathbf{x}_J, \mathbf{q}_J) \\ &\geq \min_{\emptyset \neq J \subset \{1,...,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq f(X_n) \end{aligned} \tag{6.2}$$

where $X_n := \sum_{i=1}^n x_i$.

The proof follows by Theorem 4.1 and the details are omitted.

We observe that, for a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, a sufficient condition for the positivity of $I_f(\mathbf{x}, \mathbf{q})$ for any probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero is that $f(X_n) \geq 0$. In the scalar case and if $\mathbf{x} = \mathbf{p} \in \mathbb{P}^n$, then a sufficient condition for the positivity of the f-divergence $I_f(\mathbf{p}, \mathbf{q})$ is that $f(1) \geq 0$.

The case of functions of a real variable that is of interest for applications is incorporated in:

COROLLARY 6.2 (Dragomir, 2010 [**55**]). *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function. Then for any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ we have*

$$I_f(\mathbf{p}, \mathbf{q}) \geq \max_{\emptyset \neq J \subset \{1,...,n\}} \left[Q_J f\left(\frac{P_J}{Q_J}\right) + (1 - Q_J) f\left(\frac{1 - P_J}{1 - Q_J}\right)\right] (\geq 0). \tag{6.3}$$

In what follows we provide some lower bounds for a number of f-divergences that are used in various fields of Information Theory, Probability Theory and Statistics.

The *total variation distance* is defined by the convex function $f(t) = |t - 1|, t \in \mathbb{R}$ and given in:

$$V(p, q) := \sum_{j=1}^n q_j \left|\frac{p_j}{q_j} - 1\right| = \sum_{j=1}^n |p_j - q_j|. \tag{6.4}$$

The following improvement of the positivity inequality for the total variation distance can be stated as follows.

PROPOSITION 6.3. *For any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, we have the inequality:*

$$V(p, q) \geq 2 \max_{\emptyset \neq J \subset \{1,...,n\}} |P_J - Q_J| \quad (\geq 0). \tag{6.5}$$

The proof follows by the inequality (6.3) for $f(t) = |t - 1|, t \in \mathbb{R}$.

The K. Pearson χ^2-*divergence* is obtained for the convex function $f(t) = (1 - t)^2, t \in \mathbb{R}$ and given by

$$\chi^2(p, q) := \sum_{j=1}^n q_j \left(\frac{p_j}{q_j} - 1\right)^2 = \sum_{j=1}^n \frac{(p_j - q_j)^2}{q_j}. \tag{6.6}$$

PROPOSITION 6.4. *For any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$,*

$$\begin{aligned} \chi^2(p, q) &\geq \max_{\emptyset \neq J \subset \{1,...,n\}} \left\{\frac{(P_J - Q_J)^2}{Q_J(1 - Q_J)}\right\} \\ &\geq 4 \max_{\emptyset \neq J \subset \{1,...,n\}} (P_J - Q_J)^2 \quad (\geq 0). \end{aligned} \tag{6.7}$$

PROOF. On applying the inequality (6.3) for the function $f(t) = (1-t)^2$, $t \in \mathbb{R}$, we get

$$\chi^2(p,q) \geq \max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ (1-Q_J)\left(\frac{1-P_J}{1-Q_J} - 1\right)^2 + Q_J\left(\frac{P_J}{Q_J} - 1\right)^2 \right\}$$
$$= \max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \frac{(P_J - Q_J)^2}{Q_J(1-Q_J)} \right\}.$$

Since

$$Q_J(1-Q_J) \leq \frac{1}{4}\left[Q_J + (1-Q_J)\right]^2 = \frac{1}{4},$$

then

$$\frac{(P_J - Q_J)^2}{Q_J(1-Q_J)} \geq 4(P_J - Q_J)^2$$

for each $J \subset \{1,\dots,n\}$, which proves the last part of (6.7). ∎

The *Kullback-Leibler divergence* can be obtained for the convex function $f : (0,\infty) \to \mathbb{R}$, $f(t) = t \ln t$ and is defined by

$$KL(p,q) := \sum_{j=1}^{n} q_j \cdot \frac{p_j}{q_j} \ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n} p_j \ln\left(\frac{p_j}{q_j}\right). \tag{6.8}$$

PROPOSITION 6.5. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, *we have:*

$$KL(p,q) \geq \ln\left[\max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \left(\frac{1-P_J}{1-Q_J}\right)^{1-P_J} \cdot \left(\frac{P_J}{Q_J}\right)^{P_J} \right\}\right] \geq 0. \tag{6.9}$$

PROOF. The first inequality is obvious by Corollary 6.2. Utilising the inequality between the *geometric mean and the harmonic mean,*

$$x^{\alpha} y^{1-\alpha} \geq \frac{1}{\frac{\alpha}{x} + \frac{1-\alpha}{y}}, \qquad x, y > 0,\ \alpha \in [0,1]$$

we have for $x = \frac{P_J}{Q_J}, y = \frac{1-P_J}{1-Q_J}$ and $\alpha = P_J$ that

$$\left(\frac{1-P_J}{1-Q_J}\right)^{1-P_J} \cdot \left(\frac{P_J}{Q_J}\right)^{P_J} \geq 1,$$

for any $J \subset \{1,\dots,n\}$, which implies the second part of (6.9). ∎

Another divergence measure that is of importance in Information Theory is the *Jeffreys divergence*

$$J(p,q) := \sum_{j=1}^{n} q_j \cdot \left(\frac{p_j}{q_j} - 1\right) \ln\left(\frac{p_j}{q_j}\right) = \sum_{j=1}^{n} (p_j - q_j) \ln\left(\frac{p_j}{q_j}\right), \tag{6.10}$$

which is an f-divergence for $f(t) = (t-1)\ln t$, $t > 0$.

PROPOSITION 6.6. *For any* $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$, *we have:*

$$J(p,q) \geq \ln\left(\max_{\emptyset \neq J \subset \{1,\dots,n\}} \left\{ \left[\frac{(1-P_J)Q_J}{(1-Q_J)P_J}\right]^{(Q_J - P_J)} \right\}\right) \tag{6.11}$$
$$\geq \max_{\emptyset \neq J \subset \{1,\dots,n\}} \left[\frac{(Q_J - P_J)^2}{P_J + Q_J - 2P_J Q_J}\right] \geq 0.$$

PROOF. On making use of the inequality (6.3) for $f(t) = (t-1)\ln t$, we have

$$J(p,q) \geq \max_{k\in\{1,\dots,n\}} \left\{ (1-Q_J)\left[\left(\frac{1-P_J}{1-Q_J}-1\right)\ln\left(\frac{1-P_J}{1-Q_J}\right)\right] + Q_J\left(\frac{P_J}{Q_J}-1\right)\ln\left(\frac{P_J}{Q_J}\right)\right\}$$
$$= \max_{k\in\{1,\dots,n\}} \left\{ (Q_J-P_J)\ln\left(\frac{1-P_J}{1-Q_J}\right) - (Q_J-P_J)\ln\left(\frac{P_J}{Q_J}\right)\right\}$$
$$= \max_{k\in\{1,\dots,n\}} \left\{ (Q_J-P_J)\ln\left[\frac{(1-P_J)\,Q_J}{(1-Q_J)\,P_J}\right]\right\},$$

proving the first inequality in (6.11).

Utilising the elementary inequality for positive numbers,

$$\frac{\ln b - \ln a}{b-a} \geq \frac{2}{a+b}, \qquad a,b>0$$

we have

$$(Q_J-P_J)\left[\ln\left(\frac{1-P_J}{1-Q_J}\right) - \ln\left(\frac{P_J}{Q_J}\right)\right]$$
$$= (Q_J-P_J)\cdot\frac{\ln\left(\frac{1-P_J}{1-Q_J}\right)-\ln\left(\frac{P_J}{Q_J}\right)}{\frac{1-P_J}{1-Q_J}-\frac{P_J}{Q_J}}\cdot\left[\frac{1-P_J}{1-Q_J}-\frac{P_J}{Q_J}\right]$$
$$= \frac{(Q_J-P_J)^2}{Q_J(1-Q_J)}\cdot\frac{\ln\left(\frac{1-P_J}{1-Q_J}\right)-\ln\left(\frac{P_J}{Q_J}\right)}{\frac{1-P_J}{1-Q_J}-\frac{P_J}{Q_J}}$$
$$\geq \frac{(Q_J-P_J)^2}{Q_J(1-Q_J)}\cdot\frac{2}{\frac{1-P_J}{1-Q_J}+\frac{P_J}{Q_J}} = \frac{2(Q_J-P_J)^2}{P_J+Q_J-2P_JQ_J} \geq 0,$$

for each $J \subset \{1,\dots,n\}$, giving the second inequality in (6.11). ∎

CHAPTER 10

Inequalities in Terms of Gâteaux Derivatives

1. Gâteaux Derivatives

Assume that $f : X \to \mathbb{R}$ is a *convex function* on the real linear space X. Since for any vectors $x, y \in X$ the function $g_{x,y} : \mathbb{R} \to \mathbb{R}$, $g_{x,y}(t) := f(x + ty)$ is convex it follows that the following limits exist

$$\nabla_{+(-)} f(x)(y) := \lim_{t \to 0+(-)} \frac{f(x + ty) - f(x)}{t}$$

and they are called the *right(left) Gâteaux derivatives* of the function f in the point x over the direction y.

It is obvious that for any $t > 0 > s$ we have

$$\frac{f(x + ty) - f(x)}{t} \geq \nabla_+ f(x)(y) = \inf_{t>0} \left[\frac{f(x + ty) - f(x)}{t} \right] \geq \sup_{s<0} \left[\frac{f(x + sy) - f(x)}{s} \right] = \nabla_- f(x)(y) \geq \frac{f(x + sy) - f(x)}{s} \tag{1.1}$$

for any $x, y \in X$ and, in particular,

$$\nabla_- f(u)(u - v) \geq f(u) - f(v) \geq \nabla_+ f(v)(u - v) \tag{1.2}$$

for any $u, v \in X$. We call this *the gradient inequality* for the convex function f. It will be used frequently in the sequel in order to obtain various results related to Jensen's inequality.

The following properties are also of importance:

$$\nabla_+ f(x)(-y) = -\nabla_- f(x)(y), \tag{1.3}$$

and

$$\nabla_{+(-)} f(x)(\alpha y) = \alpha \nabla_{+(-)} f(x)(y) \tag{1.4}$$

for any $x, y \in X$ and $\alpha \geq 0$.

The right Gâteaux derivative is *subadditive* while the left one is *superadditive*, i.e.,

$$\nabla_+ f(x)(y + z) \leq \nabla_+ f(x)(y) + \nabla_+ f(x)(z) \tag{1.5}$$

and

$$\nabla_- f(x)(y + z) \geq \nabla_- f(x)(y) + \nabla_- f(x)(z) \tag{1.6}$$

for any $x, y, z \in X$.

Some natural examples can be provided by the use of normed spaces.

Assume that $(X, \|\cdot\|)$ is a real normed linear space. The function $f : X \to \mathbb{R}$, $f(x) := \frac{1}{2}\|x\|^2$ is a convex function which generates *the superior* and *the inferior semi-inner products*

$$\langle y, x \rangle_{s(i)} := \lim_{t \to 0+(-)} \frac{\|x + ty\|^2 - \|x\|^2}{t}.$$

For a comprehensive study of the properties of these mappings in the Geometry of Banach Spaces see the monograph [**49**].

For the convex function $f_p : X \to \mathbb{R}$, $f_p(x) := \|x\|^p$ with $p > 1$, we have

$$\nabla_{+(-)} f_p(x)(y) = \begin{cases} p\|x\|^{p-2} \langle y, x\rangle_{s(i)} & \text{if } x \neq 0 \\ \\ 0 & \text{if } x = 0 \end{cases}$$

for any $y \in X$.

If $p = 1$, then we have

$$\nabla_{+(-)} f_1(x)(y) = \begin{cases} \|x\|^{-1} \langle y, x\rangle_{s(i)} & \text{if } x \neq 0 \\ \\ +(-)\|y\| & \text{if } x = 0 \end{cases}$$

for any $y \in X$.

This class of functions will be used to illustrate the inequalities obtained in the general case of convex functions defined on an entire linear space.

The following result holds:

THEOREM 1.1 (Dragomir, 2011 [**57**]). *Let $f : X \to \mathbb{R}$ be a convex function. Then for any $x, y \in X$ and $t \in [0, 1]$ we have*

$$\begin{aligned} &t(1-t)[\nabla_- f(y)(y-x) - \nabla_+ f(x)(y-x)] \\ &\qquad \geq tf(x) + (1-t)f(y) - f(tx + (1-t)y) \\ &\qquad \geq t(1-t)[\nabla_+ f(tx + (1-t)y)(y-x) - \nabla_- f(tx + (1-t)y)(y-x)] \geq 0. \end{aligned} \tag{1.7}$$

PROOF. Utilising the gradient inequality (1.2) we have

$$f(tx + (1-t)y) - f(x) \geq (1-t)\nabla_+ f(x)(y-x) \tag{1.8}$$

and

$$f(tx + (1-t)y) - f(y) \geq -t\nabla_- f(y)(y-x). \tag{1.9}$$

If we multiply (1.8) with t and (1.9) with $1-t$ and add the resultant inequalities we obtain

$$\begin{aligned} &f(tx + (1-t)y) - tf(x) - (1-t)f(y) \\ &\qquad \geq (1-t)t\nabla_+ f(x)(y-x) - t(1-t)\nabla_- f(y)(y-x) \end{aligned}$$

which is clearly equivalent with the first part of (1.7).

By the gradient inequality we also have

$$(1-t)\nabla_- f(tx + (1-t)y)(y-x) \geq f(tx + (1-t)y) - f(x)$$

and

$$-t\nabla_+ f(tx + (1-t)y)(y-x) \geq f(tx + (1-t)y) - f(y)$$

which by the same procedure as above yields the second part of (1.7). ∎

The following particular case for norms may be stated:

COROLLARY 1.2 (Dragomir, 2011 [**57**]). *If x and y are two vectors in the normed linear space $(X, \|\cdot\|)$ such that $0 \notin [x, y] := \{(1-s)x + sy, s \in [0,1]\}$, then for any $p \geq 1$ we have the inequalities*

$$\begin{aligned} &pt(1-t)\left[\|y\|^{p-2}\langle y-x, y\rangle_i - \|x\|^{p-2}\langle y-x, x\rangle_s\right] \\ &\qquad \geq t\|x\|^p + (1-t)\|y\|^p - \|tx + (1-t)y\|^p \\ &\qquad \geq pt(1-t)\|tx + (1-t)y\|^{p-2}\left[\langle y-x, tx + (1-t)y\rangle_s - \langle y-x, tx + (1-t)y\rangle_i\right] \geq 0 \end{aligned} \tag{1.10}$$

for any $t \in [0, 1]$. If $p \geq 2$ the inequality holds for any x and y.

REMARK 1.1. We observe that for $p = 1$ in (1.10) we derive the result

$$t(1-t)\left[\left\langle y-x, \frac{y}{\|y\|}\right\rangle_i - \left\langle y-x, \frac{x}{\|x\|}\right\rangle_s\right] \geq t\|x\| + (1-t)\|y\| - \|tx+(1-t)y\| \geq t(1-t)\left[\left\langle y-x, \frac{tx+(1-t)y}{\|tx+(1-t)y\|}\right\rangle_s - \left\langle y-x, \frac{tx+(1-t)y}{\|tx+(1-t)y\|}\right\rangle_i\right] \geq 0 \tag{1.11}$$

while for $p = 2$ we have

$$2t(1-t)\left[\langle y-x, y\rangle_i - \langle y-x, x\rangle_s\right] \geq t\|x\|^2 + (1-t)\|y\|^2 - \|tx+(1-t)y\|^2 \geq 2t(1-t)\left[\langle y-x, tx+(1-t)y\rangle_s - \langle y-x, tx+(1-t)y\rangle_i\right] \geq 0. \tag{1.12}$$

We notice that the inequality (1.12) holds for any $x, y \in X$ while in the inequality (1.11) we must assume that x, y and $tx + (1-t)y$ are not zero.

REMARK 1.2. If the normed space is smooth, i.e., the norm is Gâteaux differentiable in any nonzero point, then the superior and inferior semi-inner products coincide with the Lumer-Giles semi-inner product $[\cdot,\cdot]$ that generates the norm and is linear in the first variable (see for instance [**49**]). In this situation the inequality (1.10) becomes

$$pt(1-t)\left(\|y\|^{p-2}[y-x, y] - \|x\|^{p-2}[y-x, x]\right) \geq t\|x\|^p + (1-t)\|y\|^p - \|tx+(1-t)y\|^p \geq 0 \tag{1.13}$$

and holds for any nonzero x and y.

Moreover, if $(X, \langle\cdot,\cdot\rangle)$ is an inner product space, then (1.13) becomes

$$pt(1-t)\left\langle y-x, \|y\|^{p-2}y - \|x\|^{p-2}x\right\rangle \geq t\|x\|^p + (1-t)\|y\|^p - \|tx+(1-t)y\|^p \geq 0. \tag{1.14}$$

From (1.14) we deduce the particular inequalities of interest

$$t(1-t)\left\langle y-x, \frac{y}{\|y\|} - \frac{x}{\|x\|}\right\rangle \geq t\|x\| + (1-t)\|y\| - \|tx+(1-t)y\| \geq 0 \tag{1.15}$$

and

$$2t(1-t)\|y-x\|^2 \geq t\|x\|^2 + (1-t)\|y\|^2 - \|tx+(1-t)y\|^2 \geq 0. \tag{1.16}$$

Obviously, the inequality (1.16) can be proved directly on utilising the properties of the inner products.

2. A REFINEMENT OF JENSEN'S INEQUALITY

For a convex function $f : X \to \mathbb{R}$ defined on a linear space X, perhaps one of the most important result is the well known Jensen's inequality

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i f(x_i), \tag{2.1}$$

which holds for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$.

The following refinement of Jensen's inequality holds:

THEOREM 2.1 (Dragomir, 2011 [**57**]). *Let $f : X \to \mathbb{R}$ be a convex function defined on a linear space X. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ we have the inequality*

$$\sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \geq \sum_{k=1}^{n} p_k \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)(x_k) - \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(\sum_{i=1}^{n} p_i x_i\right) \geq 0. \tag{2.2}$$

In particular, for the uniform distribution, we have

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \geq \frac{1}{n}\left[\sum_{k=1}^{n} \nabla_+ f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)(x_k) - \nabla_+ f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)\right] \geq 0. \tag{2.3}$$

PROOF. Utilising the gradient inequality (1.2) we have

$$f(x_k) - f\left(\sum_{i=1}^{n} p_i x_i\right) \geq \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(x_k - \sum_{i=1}^{n} p_i x_i\right) \tag{2.4}$$

for any $k \in \{1, ..., n\}$.

By the subadditivity of the functional $\nabla_+ f(\cdot)(\cdot)$ in the second variable we also have

$$\nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(x_k - \sum_{i=1}^{n} p_i x_i\right) \geq \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)(x_k) - \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(\sum_{i=1}^{n} p_i x_i\right) \tag{2.5}$$

for any $k \in \{1, ..., n\}$.

Utilising the inequalities (2.4) and (2.5) we get

$$f(x_k) - f\left(\sum_{i=1}^{n} p_i x_i\right) \geq \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)(x_k) - \nabla_+ f\left(\sum_{i=1}^{n} p_i x_i\right)\left(\sum_{i=1}^{n} p_i x_i\right) \tag{2.6}$$

for any $k \in \{1, ..., n\}$.

Now, if we multiply (2.6) with $p_k \geq 0$ and sum over k from 1 to n, then we deduce the first inequality in (2.2). The second inequality is obvious by the subadditivity property of the functional $\nabla_+ f(\cdot)(\cdot)$ in the second variable. ∎

The following particular case that provides a refinement for the generalised triangle inequality in normed linear spaces is of interest

COROLLARY 2.2 (Dragomir, 2011 [**57**]). *Let $(X, \|\cdot\|)$ be a normed linear space. Then for any $p \geq 1$, for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and any probability distribution*

$\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ *with* $\sum_{i=1}^n p_i x_i \neq 0$ *we have the inequality*

$$\sum_{i=1}^n p_i \|x_i\|^p - \left\|\sum_{i=1}^n p_i x_i\right\|^p \geq p \left\|\sum_{i=1}^n p_i x_i\right\|^{p-2} \left[\sum_{k=1}^n p_k \left\langle x_k, \sum_{j=1}^n p_j x_j \right\rangle_s - \left\|\sum_{i=1}^n p_i x_i\right\|^2\right] \geq 0. \tag{2.7}$$

If $p \geq 2$ *the inequality holds for any* n*-tuple of vectors and probability distribution.*

In particular, we have the norm inequalities

$$\sum_{i=1}^n p_i \|x_i\| - \left\|\sum_{i=1}^n p_i x_i\right\| \geq \left[\sum_{k=1}^n p_k \left\langle x_k, \frac{\sum_{i=1}^n p_i x_i}{\|\sum_{i=1}^n p_i x_i\|} \right\rangle_s - \left\|\sum_{i=1}^n p_i x_i\right\|\right] \geq 0. \tag{2.8}$$

and

$$\sum_{i=1}^n p_i \|x_i\|^2 - \left\|\sum_{i=1}^n p_i x_i\right\|^2 \geq 2 \left[\sum_{k=1}^n p_k \left\langle x_k, \sum_{i=1}^n p_i x_i \right\rangle_s - \left\|\sum_{i=1}^n p_i x_i\right\|^2\right] \geq 0. \tag{2.9}$$

We notice that the first inequality in (2.9) is equivalent with

$$\sum_{i=1}^n p_i \|x_i\|^2 + \left\|\sum_{i=1}^n p_i x_i\right\|^2 \geq 2 \sum_{k=1}^n p_k \left\langle x_k, \sum_{i=1}^n p_i x_i \right\rangle_s$$

which provides the result

$$\frac{1}{2}\left[\sum_{i=1}^n p_i \|x_i\|^2 + \left\|\sum_{i=1}^n p_i x_i\right\|^2\right] \geq \sum_{k=1}^n p_k \left\langle x_k, \sum_{i=1}^n p_i x_i \right\rangle_s \left(\geq \left\|\sum_{i=1}^n p_i x_i\right\|^2\right) \tag{2.10}$$

for any n-tuple of vectors and probability distribution.

REMARK 2.1. If in the inequality (2.7) we consider the uniform distribution, then we get

$$\sum_{i=1}^n \|x_i\|^p - n^{1-p} \left\|\sum_{i=1}^n x_i\right\|^p \geq p n^{1-p} \left\|\sum_{i=1}^n x_i\right\|^{p-2} \left[\sum_{k=1}^n \left\langle x_k, \sum_{i=1}^n x_i \right\rangle_s - \left\|\sum_{i=1}^n x_i\right\|^2\right] \geq 0. \tag{2.11}$$

3. A REVERSE OF JENSEN'S INEQUALITY

The following result is of interest as well:

THEOREM 3.1 (Dragomir, 2011 [**57**]). *Let $f : X \to \mathbb{R}$ be a convex function defined on a linear space X. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ we have the inequality*

$$\sum_{k=1}^{n} p_k \nabla_- f(x_k)(x_k) - \sum_{k=1}^{n} p_k \nabla_- f(x_k)\left(\sum_{i=1}^{n} p_i x_i\right) \geq \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right). \tag{3.1}$$

In particular, for the uniform distribution, we have

$$\frac{1}{n}\left[\sum_{k=1}^{n} \nabla_- f(x_k)(x_k) - \sum_{k=1}^{n} \nabla_- f(x_k)\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\right] \geq \frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right). \tag{3.2}$$

PROOF. Utilising the gradient inequality (1.2) we can state that

$$\nabla_- f(x_k)\left(x_k - \sum_{i=1}^{n} p_i x_i\right) \geq f(x_k) - f\left(\sum_{i=1}^{n} p_i x_i\right) \tag{3.3}$$

for any $k \in \{1, ..., n\}$.

By the superadditivity of the functional $\nabla_- f(\cdot)(\cdot)$ in the second variable we also have

$$\nabla_- f(x_k)(x_k) - \nabla_- f(x_k)\left(\sum_{i=1}^{n} p_i x_i\right) \geq \nabla_- f(x_k)\left(x_k - \sum_{i=1}^{n} p_i x_i\right) \tag{3.4}$$

for any $k \in \{1, ..., n\}$.

Therefore, by (3.3) and (3.4) we get

$$\nabla_- f(x_k)(x_k) - \nabla_- f(x_k)\left(\sum_{i=1}^{n} p_i x_i\right) \geq f(x_k) - f\left(\sum_{i=1}^{n} p_i x_i\right) \tag{3.5}$$

for any $k \in \{1, ..., n\}$.

Finally, by multiplying (3.5) with $p_k \geq 0$ and summing over k from 1 to n we deduce the desired inequality (3.1). ∎

REMARK 3.1. If the function f is defined on the Euclidian space $\mathbb{R}^n$ and is differentiable and convex, then from (3.1) we get the inequality

$$\sum_{k=1}^{n} p_k \langle \nabla f(x_k), x_k \rangle - \left\langle \sum_{k=1}^{n} p_k \nabla f(x_k), \sum_{i=1}^{n} p_i x_i \right\rangle \geq \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \tag{3.6}$$

where, as usual, for $x_k = (x_k^1, ..., x_k^n)$, $\nabla f(x_k) = \left(\frac{\partial f(x_k)}{\partial x^1}, ..., \frac{\partial f(x_k)}{\partial x^n}\right)$. This inequality was obtained firstly by Dragomir & Goh in 1996, see [**70**].

For one dimension we get the inequality

$$\sum_{k=1}^{n} p_k x_k f'(x_k) - \sum_{i=1}^{n} p_i x_i \sum_{k=1}^{n} p_k f'(x_k) \geq \sum_{i=1}^{n} p_i f(x_i) - f\left(\sum_{i=1}^{n} p_i x_i\right) \tag{3.7}$$

that was discovered in 1994 by Dragomir and Ionescu, see [**62**].

The following reverse of the generalised triangle inequality holds:

COROLLARY 3.2 (Dragomir, 2011 [**57**]). *Let $(X, \|\cdot\|)$ be a normed linear space. Then for any $p \geq 1$, for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n \setminus \{(0, ..., 0)\}$ and any probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ we have the inequality*

$$p\left[\sum_{k=1}^{n} p_k \|x_k\|^p - \sum_{k=1}^{n} p_k \|x_k\|^{p-2} \left\langle \sum_{i=1}^{n} p_i x_i, x_k \right\rangle_i\right] \geq \sum_{i=1}^{n} p_i \|x_i\|^p - \left\|\sum_{i=1}^{n} p_i x_i\right\|^p. \tag{3.8}$$

In particular, we have the norm inequalities

$$\sum_{k=1}^{n} p_k \|x_k\| - \sum_{k=1}^{n} p_k \left\langle \sum_{i=1}^{n} p_i x_i, \frac{x_k}{\|x_k\|} \right\rangle_i \geq \sum_{i=1}^{n} p_i \|x_i\| - \left\|\sum_{i=1}^{n} p_i x_i\right\| \tag{3.9}$$

for $x_k \neq 0, k \in \{1, ..., n\}$ and

$$2\left[\sum_{k=1}^{n} p_k \|x_k\|^2 - \sum_{k=1}^{n} p_k \left\langle \sum_{j=1}^{n} p_j x_j, x_k \right\rangle_i\right] \geq \sum_{i=1}^{n} p_i \|x_i\|^2 - \left\|\sum_{i=1}^{n} p_i x_i\right\|^2, \tag{3.10}$$

for any x_k.

We observe that the inequality (3.10) is equivalent with

$$\sum_{i=1}^{n} p_i \|x_i\|^2 + \left\|\sum_{i=1}^{n} p_i x_i\right\|^2 \geq 2\sum_{k=1}^{n} p_k \left\langle \sum_{j=1}^{n} p_j x_j, x_k \right\rangle_i$$

which provides the interesting result

$$\frac{1}{2}\left[\sum_{i=1}^{n} p_i \|x_i\|^2 + \left\|\sum_{i=1}^{n} p_i x_i\right\|^2\right] \geq \sum_{k=1}^{n} p_k \left\langle \sum_{j=1}^{n} p_j x_j, x_k \right\rangle_i \left(\geq \sum_{k=1}^{n}\sum_{j=1}^{n} p_j p_k \langle x_j, x_k \rangle_i\right) \tag{3.11}$$

holding for any n-tuple of vectors and probability distribution.

REMARK 3.2. If in the inequality (3.8) we consider the uniform distribution, then we get

$$p\left[\sum_{k=1}^{n}\|x_k\|^p-\frac{1}{n}\sum_{k=1}^{n}\|x_k\|^{p-2}\left\langle\sum_{j=1}^{n}x_j,x_k\right\rangle_i\right] \geq \sum_{i=1}^{n}\|x_i\|^p-n^{1-p}\left\|\sum_{i=1}^{n}x_i\right\|^p. \tag{3.12}$$

For $p\in[1,2)$ all vectors x_k should not be zero.

4. BOUNDS FOR THE MEAN f-DEVIATION

Let X be a real linear space. For a convex function $f:X\to\mathbb{R}$ with the property that $f(0)\geq 0$ we define the *mean f-deviation* of an n-tuple of vectors $y=(y_1,...,y_n)\in X^n$ with the probability distribution $\mathbf{p}=(p_1,...,p_n)\in\mathbb{P}^n$ by the non-negative quantity

$$K_{f(\cdot)}(\mathbf{p},\mathbf{y})=K_f(\mathbf{p},\mathbf{y}):=\sum_{i=1}^{n}p_if\left(y_i-\sum_{k=1}^{n}p_ky_k\right). \tag{4.1}$$

The fact that $K_f(\mathbf{p},\mathbf{y})$ is non-negative follows by Jensen's inequality, namely

$$K_f(\mathbf{p},\mathbf{y})\geq f\left(\sum_{i=1}^{n}p_i\left(y_i-\sum_{k=1}^{n}p_ky_k\right)\right)=f(0)\geq 0.$$

Of course the concept can be extended for any function defined on $X,$ however if the function is not convex or if it is convex but $f(0)<0,$ then we are not sure about the positivity of the quantity $K_f(\mathbf{p},\mathbf{y})$.

A natural example of such deviations can be provided by the convex function $f(y):=\|y\|^r$ with $r\geq 1$ defined on a normed linear space $(X,\|\cdot\|)$. We denote this by

$$K_r(\mathbf{p},\mathbf{y}):=\sum_{i=1}^{n}p_i\left\|y_i-\sum_{k=1}^{n}p_ky_k\right\|^r \tag{4.2}$$

and call it the *mean r-absolute deviation* of the n-tuple of vectors $y=(y_1,...,y_n)\in X^n$ with the probability distribution $\mathbf{p}=(p_1,...,p_n)$.

Utilising the result from [**54**] we can state then the following result providing a non-trivial lower bound for the mean f-deviation:

THEOREM 4.1. *Let $f:X\to[0,\infty)$ be a convex function with $f(0)=0.$ If $y=(y_1,...,y_n)\in X^n$ and $\mathbf{p}=(p_1,...,p_n)$ is a probability distribution with all p_i nonzero, then*

$$K_f(\mathbf{p},\mathbf{y}) \geq \max_{k\in\{1,...,n\}}\left\{(1-p_k)f\left[\frac{p_k}{1-p_k}\left(y_k-\sum_{l=1}^{n}p_ly_l\right)\right]+p_kf\left(y_k-\sum_{l=1}^{n}p_ly_l\right)\right\}(\geq 0). \tag{4.3}$$

The case for mean r-absolute deviation is incorporated in

COROLLARY 4.2. *Let $(X,\|\cdot\|)$ be a normed linear space. If $y=(y_1,...,y_n)\in X^n$ and $\mathbf{p}=(p_1,...,p_n)$ is a probability distribution with all p_i nonzero, then for $r\geq 1$ we have*

$$K_r(\mathbf{p},\mathbf{y})\geq\max_{k\in\{1,...,n\}}\left\{\left[(1-p_k)^{1-r}p_k^r+p_k\right]\left\|y_k-\sum_{l=1}^{n}p_ly_l\right\|^r\right\}. \tag{4.4}$$

REMARK 4.1. Since the function $h_r(t) := (1-t)^{1-r} t^r + t$, $r \geq 1$, $t \in [0,1)$ is strictly increasing on $[0,1)$, then

$$\min_{k\in\{1,\ldots,n\}} \left\{(1-p_k)^{1-r} p_k^r + p_k\right\} = p_m + (1-p_m)^{1-r} p_m^r,$$

where $p_m := \min_{k\in\{1,\ldots,n\}} p_k$. We then obtain the following simpler inequality:

$$K_r(\mathbf{p},\mathbf{y}) \geq \left[p_m + (1-p_m)^{1-r} \cdot p_m^r\right] \max_{k\in\{1,\ldots,n\}} \left\|y_k - \sum_{l=1}^n p_l y_l\right\|^p, \tag{4.5}$$

which is perhaps more useful for applications.

We have the following double inequality for the mean f-mean deviation:

THEOREM 4.3 (Dragomir, 2011 [**57**]). *Let $f : X \to [0,\infty)$ be a convex function with $f(0) = 0$. If $y = (y_1, \ldots, y_n) \in X^n$ and $\mathbf{p} = (p_1, \ldots, p_n)$ is a probability distribution with all p_i nonzero, then*

$$K_{\nabla_- f(\cdot)(\cdot)}(\mathbf{p},\mathbf{y}) \geq K_{f(\cdot)}(\mathbf{p},\mathbf{y}) \geq K_{\nabla_+ f(0)(\cdot)}(\mathbf{p},\mathbf{y}) \geq 0. \tag{4.6}$$

PROOF. If we use the inequality (2.2) for $x_i = y_i - \sum_{k=1}^n p_k y_k$ we get

$$\begin{aligned}
&\sum_{i=1}^n p_i f\left(y_i - \sum_{k=1}^n p_k y_k\right) - f\left(\sum_{i=1}^n p_i\left(y_i - \sum_{k=1}^n p_k y_k\right)\right)\\
&\qquad \geq \sum_{j=1}^n p_j \nabla_+ f\left(\sum_{i=1}^n p_i\left(y_i - \sum_{k=1}^n p_k y_k\right)\right)\left(y_j - \sum_{k=1}^n p_k y_k\right)\\
&\qquad\qquad - \nabla_+ f\left(\sum_{i=1}^n p_i\left(y_i - \sum_{k=1}^n p_k y_k\right)\right)\left(\sum_{i=1}^n p_i\left(y_i - \sum_{k=1}^n p_k y_k\right)\right) \geq 0
\end{aligned}$$

which is equivalent with the second part of (4.6).

Now, by utilising the inequality (3.1) for the same choice of x_i we get

$$\begin{aligned}
&\sum_{j=1}^n p_j \nabla_- f\left(y_j - \sum_{k=1}^n p_k y_k\right)\left(y_j - \sum_{k=1}^n p_k y_k\right)\\
&\qquad - \sum_{k=1}^n p_j \nabla_- f\left(y_j - \sum_{k=1}^n p_k y_k\right)\left(\sum_{i=1}^n p_i\left(y_i - \sum_{k=1}^n p_k y_k\right)\right)\\
&\qquad\qquad \geq \sum_{i=1}^n p_i f\left(y_i - \sum_{k=1}^n p_k y_k\right) - f\left(\sum_{i=1}^n p_i\left(y_i - \sum_{k=1}^n p_k y_k\right)\right),
\end{aligned}$$

which in its turn is equivalent with the first inequality in (4.6). ∎

We observe that as examples of convex functions defined on the entire normed linear space $(X, \|\cdot\|)$ that are convex and vanishes in 0 we can consider the functions

$$f(x) := g(\|x\|), \ x \in X$$

where $g : [0,\infty) \to [0,\infty)$ is a monotonic nondecreasing convex function with $g(0) = 0$.

For this kind of functions we have by direct computation that

$$\nabla_+ f(0)(u) = g'_+(0)\|u\| \ \text{ for any } u \in X$$

and

$$\nabla_{-} f(u)(u) = g'_{-}(\|u\|)\|u\| \text{ for any } u \in X.$$

We then have the following norm inequalities that are of interest:

COROLLARY 4.4 (Dragomir, 2011 [**57**]). *Let* $(X, \|\cdot\|)$ *be a normed linear space. If* $g : [0,\infty) \to [0,\infty)$ *is a monotonic nondecreasing convex function with* $g(0) = 0$, *then for any* $y = (y_1, ..., y_n) \in X^n$ *and* $\mathbf{p} = (p_1, ..., p_n)$ *a probability distribution, we have*

$$\sum_{i=1}^{n} p_i g'_{-}\left(\left\|y_i - \sum_{k=1}^{n} p_k y_k\right\|\right)\left\|y_i - \sum_{k=1}^{n} p_k y_k\right\| \geq \sum_{i=1}^{n} p_i g\left(\left\|y_i - \sum_{k=1}^{n} p_k y_k\right\|\right) \geq g'_{+}(0)\sum_{i=1}^{n} p_i\left\|y_i - \sum_{k=1}^{n} p_k y_k\right\|. \tag{4.7}$$

5. BOUNDS FOR f-DIVERGENCE MEASURES

We endeavour to extend this concept for functions defined on a cone in a linear space as follows (see also [**55**]).

Firstly, we recall that the subset K in a linear space X is a *cone* if the following two conditions are satisfied:

(i) for any $x, y \in K$ we have $x + y \in K$;

(ii) for any $x \in K$ and any $\alpha \geq 0$ we have $\alpha x \in K$.

For a given n-tuple of vectors $\mathbf{z} = (z_1, ..., z_n) \in K^n$ and a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero, we can define, for the convex function $f : K \to \mathbb{R}$, the following *f-divergence of* $\mathbf{z}$ *with the distribution* $\mathbf{q}$

$$I_f(\mathbf{z}, \mathbf{q}) := \sum_{i=1}^{n} q_i f\left(\frac{z_i}{q_i}\right). \tag{5.1}$$

It is obvious that if $X = \mathbb{R}$, $K = [0,\infty)$ and $\mathbf{x} = \mathbf{p} \in \mathbb{P}^n$ then we obtain the usual concept of the f-divergence associated with a function $f : [0,\infty) \to \mathbb{R}$.

Now, for a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1, ..., n\}$ we have

$$\mathbf{q}_J := \left(Q_J, \bar{Q}_J\right) \in \mathbb{P}^2$$

and

$$\mathbf{x}_J := \left(X_J, \bar{X}_J\right) \in K^2$$

where, as above

$$X_J := \sum_{i \in J} x_i, \quad \text{and} \quad \bar{X}_J := X_{\bar{J}}.$$

It is obvious that

$$I_f(\mathbf{x}_J, \mathbf{q}_J) = Q_J f\left(\frac{X_J}{Q_J}\right) + \bar{Q}_J f\left(\frac{\bar{X}_J}{\bar{Q}_J}\right).$$

The following inequality for the f-divergence of an n-tuple of vectors in a linear space holds [**55**]:

THEOREM 5.1. *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero and for any nonempty subset J of $\{1, ..., n\}$ we have*

$$I_f(\mathbf{x}, \mathbf{q}) \geq \max_{\emptyset \neq J \subset \{1,...,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq I_f(\mathbf{x}_J, \mathbf{q}_J) \geq \min_{\emptyset \neq J \subset \{1,...,n\}} I_f(\mathbf{x}_J, \mathbf{q}_J) \geq f(X_n) \tag{5.2}$$

where $X_n := \sum_{i=1}^n x_i$.

We observe that, for a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$, a sufficient condition for the positivity of $I_f(\mathbf{x}, \mathbf{q})$ for any probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero is that $f(X_n) \geq 0$. In the scalar case and if $\mathbf{x} = \mathbf{p} \in \mathbb{P}^n$, then a sufficient condition for the positivity of the f-divergence $I_f(\mathbf{p}, \mathbf{q})$ is that $f(1) \geq 0$.

The case of functions of a real variable that is of interest for applications is incorporated in [**55**]:

COROLLARY 5.2. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function. Then for any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ we have*

$$I_f(\mathbf{p}, \mathbf{q}) \geq \max_{\emptyset \neq J \subset \{1,...,n\}} \left[Q_J f\left(\frac{P_J}{Q_J}\right) + (1 - Q_J) f\left(\frac{1 - P_J}{1 - Q_J}\right)\right] (\geq 0). \tag{5.3}$$

In what follows, by the use of the results in Theorem 2.1 and Theorem 3.1 we can provide an upper and a lower bound for the positive difference $I_f(\mathbf{x}, \mathbf{q}) - f(X_n)$.

THEOREM 5.3 (Dragomir, 2011 [**57**]). *Let $f : K \to \mathbb{R}$ be a convex function on the cone K. Then for any n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in K^n$ and a probability distribution $\mathbf{q} \in \mathbb{P}^n$ with all values nonzero we have*

$$I_{\nabla_- f(\cdot)(\cdot)}(\mathbf{x}, \mathbf{q}) - I_{\nabla_- f(\cdot)(X_n)}(\mathbf{x}, \mathbf{q}) \geq I_f(\mathbf{x}, \mathbf{q}) - f(X_n) \geq I_{\nabla_+ f(X_n)(\cdot)}(\mathbf{x}, \mathbf{q}) - \nabla_+ f(X_n)(X_n) \geq 0. \tag{5.4}$$

The case of functions of a real variable that is useful for applications is as follows:

COROLLARY 5.4. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function. Then for any $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ we have*

$$I_{f'_-(\cdot)(\cdot)}(\mathbf{p}, \mathbf{q}) - I_{f'_-(\cdot)}(\mathbf{p}, \mathbf{q}) \geq I_f(\mathbf{p}, \mathbf{q}) \geq 0, \tag{5.5}$$

or, equivalently,

$$I_{f'_-(\cdot)[(\cdot)-1]}(\mathbf{p}, \mathbf{q}) \geq I_f(\mathbf{p}, \mathbf{q}) \geq 0. \tag{5.6}$$

The above corollary is useful to provide an upper bound in terms of the variational distance for the f-divergence $I_f(\mathbf{p}, \mathbf{q})$ of normalized convex functions whose derivatives are bounded above and below.

PROPOSITION 5.5. *Let $f : [0, \infty) \to \mathbb{R}$ be a normalized convex function and $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$. If there exists the constants γ and Γ with*

$$-\infty < \gamma \leq f'_-\left(\frac{p_k}{q_k}\right) \leq \Gamma < \infty \text{ for all } k \in \{1, ..., n\},$$

then we have the inequality

$$0 \leq I_f(\mathbf{p}, \mathbf{q}) \leq \frac{1}{2}(\Gamma - \gamma) V(\mathbf{p}, \mathbf{q}), \tag{5.7}$$

where $V(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^n q_i \left|\frac{p_i}{q_i} - 1\right| = \sum_{i=1}^n |p_i - q_i|$.

PROOF. By the inequality (5.6) we have successively that

$$\begin{aligned}0 &\leq I_f(\mathbf{p},\mathbf{q}) \leq I_{f'_-(\cdot)[(\cdot)-1]}(\mathbf{p},\mathbf{q}) \\ &= \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i}-1\right)\left[f'_-\left(\frac{p_i}{q_i}\right)-\frac{\Gamma+\gamma}{2}\right] \\ &\leq \sum_{i=1}^{n} q_i \left|\frac{p_i}{q_i}-1\right|\left|f'_-\left(\frac{p_i}{q_i}\right)-\frac{\Gamma+\gamma}{2}\right| \\ &\leq \frac{1}{2}(\Gamma-\gamma)\sum_{i=1}^{n} q_i\left|\frac{p_i}{q_i}-1\right|\end{aligned}$$

which proves the desired result (5.7). ∎

COROLLARY 5.6. *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function and $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$. If there exist the constants r and R with*

$$0 < r \leq \frac{p_k}{q_k} \leq R < \infty \text{ for all } k\in\{1,...,n\},$$

then we have the inequality

$$0 \leq I_f(\mathbf{p},\mathbf{q}) \leq \frac{1}{2}\left[f'_-(R)-f'_-(r)\right]V(\mathbf{p},\mathbf{q}). \tag{5.8}$$

The K. Pearson χ^2-*divergence* is obtained for the convex function $f(t)=(1-t)^2$, $t\in\mathbb{R}$ and given by

$$\chi^2(p,q) := \sum_{j=1}^{n} q_j\left(\frac{p_j}{q_j}-1\right)^2 = \sum_{j=1}^{n}\frac{(p_j-q_j)^2}{q_j}.$$

Finally, the following proposition giving another upper bound in terms of the χ^2-divergence can be stated:

PROPOSITION 5.7. *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function and $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$. If there exists the constant $0<\Delta<\infty$ with*

$$\left|\frac{f'_-\left(\frac{p_i}{q_i}\right)-f'_-(1)}{\frac{p_i}{q_i}-1}\right| \leq \Delta \text{ for all } k\in\{1,...,n\}, \tag{5.9}$$

then we have the inequality

$$0 \leq I_f(\mathbf{p},\mathbf{q}) \leq \Delta\chi^2(p,q). \tag{5.10}$$

In particular, if $f'_-(\cdot)$ satisfies the local Lipschitz condition

$$\left|f'_-(x)-f'_-(1)\right| \leq \Delta|x-1| \text{ for any } x\in(0,\infty) \tag{5.11}$$

then (5.10) holds true for any $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$.

PROOF. We have from (5.6) that

$$\begin{aligned}
0 &\leq I_f(\mathbf{p},\mathbf{q}) \leq I_{f'_-(\cdot)[(\cdot)-1]}(\mathbf{p},\mathbf{q}) \\
&= \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right)\left[f'_-\left(\frac{p_i}{q_i}\right) - f'_-(1)\right] \\
&\leq \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right)^2 \left|\frac{f'_-\left(\frac{p_i}{q_i}\right) - f'_-(1)}{\frac{p_i}{q_i} - 1}\right| \\
&\leq \Delta \sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i} - 1\right)^2
\end{aligned}$$

and the inequality (5.10) is obtained. ∎

REMARK 5.1. It is obvious that if one chooses in the above inequalities particular normalized convex functions that generates the Kullback-Leibler, Jeffreys, Hellinger or other divergence measures or discrepancies, that one can obtain some results of interest. However the details are not provided here.

CHAPTER 11

Inequalities of Slater's Type

1. Introduction

Suppose that I is an interval of real numbers with interior $\mathring{I}$ and $f : I \to \mathbb{R}$ is a convex function on I. Then f is continuous on $\mathring{I}$ and has finite left and right derivatives at each point of $\mathring{I}$. Moreover, if $x, y \in \mathring{I}$ and $x < y$, then $f'_-(x) \leq f'_+(x) \leq f'_-(y) \leq f'_+(y)$ which shows that both f'_- and f'_+ are nondecreasing function on $\mathring{I}$. It is also known that a convex function must be differentiable except for at most countably many points.

For a convex function $f : I \to \mathbb{R}$, the subdifferential of f denoted by ∂f is the set of all functions $\varphi : I \to [-\infty, \infty]$ such that $\varphi\left(\mathring{I}\right) \subset \mathbb{R}$ and

$$f(x) \geq f(a) + (x-a)\varphi(a) \text{ for any } x, a \in I.$$

It is also well known that if f is convex on I, then ∂f is nonempty, f'_-, $f'_+ \in \partial f$ and if $\varphi \in \partial f$, then

$$f'_-(x) \leq \varphi(x) \leq f'_+(x) \text{ for any } x \in \mathring{I}.$$

In particular, φ is a nondecreasing function.

If f is differentiable and convex on $\mathring{I}$, then $\partial f = \{f'\}$.

The following result is well known in the literature as *the Slater inequality:*

THEOREM 1.1 (Slater, 1981, [**123**]). *If $f : I \to \mathbb{R}$ is a nonincreasing (nondecreasing) convex function, $x_i \in I, p_i \geq 0$ with $P_n := \sum_{i=1}^n p_i > 0$ and $\sum_{i=1}^n p_i\varphi(x_i) \neq 0$, where $\varphi \in \partial f$, then*

$$\frac{1}{P_n}\sum_{i=1}^n p_i f(x_i) \leq f\left(\frac{\sum_{i=1}^n p_i x_i \varphi(x_i)}{\sum_{i=1}^n p_i \varphi(x_i)}\right). \tag{1.1}$$

As pointed out in [**48**, p. 208], the monotonicity assumption for the derivative φ can be replaced with the condition

$$\frac{\sum_{i=1}^n p_i x_i \varphi(x_i)}{\sum_{i=1}^n p_i \varphi(x_i)} \in I, \tag{1.2}$$

which is more general and can hold for suitable points in I and for not necessarily monotonic functions.

The main aim of the present paper is to extend Slater's inequality for convex functions defined on general linear spaces. A reverse of the Slater's inequality is also obtained. Natural applications for norm inequalities and f-divergence measures are provided as well.

2. Slater's Inequality for Functions Defined on Linear Spaces

Assume that $f : X \to \mathbb{R}$ is a *convex function* on the real linear space X. Since for any vectors $x, y \in X$ the function $g_{x,y} : \mathbb{R} \to \mathbb{R}$, $g_{x,y}(t) := f(x + ty)$ is convex it follows that the following limits exist

$$\nabla_{+(-)} f(x)(y) := \lim_{t \to 0+(-)} \frac{f(x+ty) - f(x)}{t}$$

and they are called the *right(left) Gâteaux derivatives* of the function f in the point x over the direction y.

It is obvious that for any $t > 0 > s$ we have

$$\frac{f(x+ty)-f(x)}{t} \geq \nabla_+ f(x)(y) = \inf_{t>0}\left[\frac{f(x+ty)-f(x)}{t}\right] \geq \sup_{s<0}\left[\frac{f(x+sy)-f(x)}{s}\right] = \nabla_- f(x)(y) \geq \frac{f(x+sy)-f(x)}{s} \tag{2.1}$$

for any $x, y \in X$ and, in particular,

$$\nabla_- f(u)(u-v) \geq f(u) - f(v) \geq \nabla_+ f(v)(u-v) \tag{2.2}$$

for any $u, v \in X$. We call this *the gradient inequality* for the convex function f. It will be used frequently in the sequel in order to obtain various results related toSlater's inequality.

The following properties are also of importance:

$$\nabla_+ f(x)(-y) = -\nabla_- f(x)(y), \tag{2.3}$$

and

$$\nabla_{+(-)} f(x)(\alpha y) = \alpha \nabla_{+(-)} f(x)(y) \tag{2.4}$$

for any $x, y \in X$ and $\alpha \geq 0$.

The right Gâteaux derivative is *subadditive* while the left one is *superadditive*, i.e.,

$$\nabla_+ f(x)(y+z) \leq \nabla_+ f(x)(y) + \nabla_+ f(x)(z) \tag{2.5}$$

and

$$\nabla_- f(x)(y+z) \geq \nabla_- f(x)(y) + \nabla_- f(x)(z) \tag{2.6}$$

for any $x, y, z \in X$.

Some natural examples can be provided by the use of normed spaces.

Assume that $(X, \|\cdot\|)$ is a real normed linear space. The function $f : X \to \mathbb{R}$, $f(x) := \frac{1}{2}\|x\|^2$ is a convex function which generates *the superior* and *the inferior semi-inner products*

$$\langle y, x\rangle_{s(i)} := \lim_{t\to 0+(-)} \frac{\|x+ty\|^2 - \|x\|^2}{t}.$$

For a comprehensive study of the properties of these mappings in the Geometry of Banach Spaces see the monograph [**49**].

For the convex function $f_p : X \to \mathbb{R}$, $f_p(x) := \|x\|^p$ with $p > 1$, we have

$$\nabla_{+(-)} f_p(x)(y) = \begin{cases} p\|x\|^{p-2}\langle y, x\rangle_{s(i)} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}$$

for any $y \in X$.

If $p = 1$, then we have

$$\nabla_{+(-)} f_1(x)(y) = \begin{cases} \|x\|^{-1}\langle y, x\rangle_{s(i)} & \text{if } x \neq 0, \\ +(-)\|y\| & \text{if } x = 0 \end{cases}$$

for any $y \in X$.

For a given convex function $f : X \to \mathbb{R}$ and a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ we consider the sets

$$Sla_{+(-)}(f, \mathbf{x}) := \left\{v \in X \mid \nabla_{+(-)} f(x_i)(v - x_i) \geq 0 \text{ for all } i \in \{1, ..., n\}\right\} \tag{2.7}$$

and

$$Sla_{+(-)}(f,\mathbf{x},\mathbf{p}) := \left\{ v \in X \mid \sum_{i=1}^{n} p_i \nabla_{+(-)} f(x_i)(v - x_i) \geq 0 \right\} \tag{2.8}$$

where $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ is a given probability distribution.

Since $\nabla_{+(-)} f(x)(0) = 0$ for any $x \in X$, then we observe that $\{x_1, ..., x_n\} \subset Sla_{+(-)}(f,\mathbf{x},\mathbf{p})$, therefore the sets $Sla_{+(-)}(f,\mathbf{x},\mathbf{p})$ are not empty for each $f, \mathbf{x}$ and $\mathbf{p}$ as above.

The following properties of these sets hold:

LEMMA 2.1 (Dragomir, 2012 [**58**]). *For a given convex function $f : X \to \mathbb{R}$, a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n$ and a given probability distribution $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ we have*

(i) $Sla_-(f,\mathbf{x}) \subset Sla_+(f,\mathbf{x})$ and $Sla_-(f,\mathbf{x},\mathbf{p}) \subset Sla_+(f,\mathbf{x},\mathbf{p})$;

(ii) $Sla_-(f,\mathbf{x}) \subset Sla_-(f,\mathbf{x},\mathbf{p})$ and $Sla_+(f,\mathbf{x}) \subset Sla_+(f,\mathbf{x},\mathbf{p})$ for all $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$;

(iii) The sets $Sla_-(f,\mathbf{x})$ and $Sla_-(f,\mathbf{x},\mathbf{p})$ are convex.

PROOF. The properties (i) and (ii) follow from the definition and the fact that $\nabla_+ f(x)(y) \geq \nabla_- f(x)(y)$ for any x, y.

(iii) Let us only prove that $Sla_-(f,\mathbf{x})$ is convex.

If we assume that $y_1, y_2 \in Sla_-(f,\mathbf{x})$ and $\alpha, \beta \in [0,1]$ with $\alpha + \beta = 1$, then by the superadditivity and positive homogeneity of the Gâteaux derivative $\nabla_- f(\cdot)(\cdot)$ in the second variable we have

$$\begin{aligned} \nabla_- f(x_i)(\alpha y_1 + \beta y_2 - x_i) &= \nabla_- f(x_i)[\alpha(y_1 - x_i) + \beta(y_2 - x_i)] \\ &\geq \alpha \nabla_- f(x_i)(y_1 - x_i) + \beta \nabla_- f(x_i)(y_2 - x_i) \geq 0 \end{aligned}$$

for all $i \in \{1, ..., n\}$, which shows that $\alpha y_1 + \beta y_2 \in Sla_-(f,\mathbf{x})$.

The proof for the convexity of $Sla_-(f,\mathbf{x},\mathbf{p})$ is similar and the details are omitted. ∎

For the convex function $f_p : X \to \mathbb{R}$, $f_p(x) := \|x\|^p$ with $p \geq 1$, defined on the normed linear space $(X, \|\cdot\|)$ and for the n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in X^n \setminus \{(0, ..., 0)\}$ we have, by the well known property of the semi inner products

$$\langle y + \alpha x, x \rangle_{s(i)} = \langle y, x \rangle_{s(i)} + \alpha \|x\|^2 \text{ for any } x, y \in X \text{ and } \alpha \in \mathbb{R},$$

that

$$\begin{aligned} Sla_{+(-)}(\|\cdot\|^p, \mathbf{x}) &= Sla_{+(-)}(\|\cdot\|, \mathbf{x}) \\ &:= \left\{ v \in X \mid \langle v, x_j \rangle_{s(i)} \geq \|x_j\|^2 \text{ for all } j \in \{1, ..., n\} \right\} \end{aligned}$$

which, as can be seen, does not depend of p. We observe that, by the continuity of the semi-inner products in the first variable that $Sla_{+(-)}(\|\cdot\|, \mathbf{x})$ is closed in $(X, \|\cdot\|)$. Also, we should remarks that if $v \in Sla_{+(-)}(\|\cdot\|, \mathbf{x})$ then for any $\gamma \geq 1$ we also have that $\gamma v \in Sla_{+(-)}(\|\cdot\|, \mathbf{x})$.

The larger classes, which are dependent on the probability distribution $\mathbf{p} \in \mathbb{P}^n$ are described by

$$Sla_{+(-)}(\|\cdot\|^p, \mathbf{x}, \mathbf{p}) := \left\{ v \in X \mid \sum_{j=1}^{n} p_j \|x_j\|^{p-2} \langle v, x_j \rangle_{s(i)} \geq \sum_{j=1}^{n} p_j \|x_j\|^p \right\}.$$

If the normed space is smooth, i.e., the norm is Gâteaux differentiable in any nonzero point, then the superior and inferior semi-inner products coincide with the Lumer-Giles semi-inner product $[\cdot,\cdot]$ that generates the norm and is linear in the first variable (see for instance [**49**]). In this situation

$$Sla\left(\left\|\cdot\right\|,\mathbf{x}\right)=\left\{v\in X \mid \left[v,x_j\right]\geq \left\|x_j\right\|^2 \text{ for all } j\in\{1,...,n\}\right\}$$

and

$$Sla\left(\left\|\cdot\right\|^p,\mathbf{x},\mathbf{p}\right)=\left\{v\in X \mid \sum_{j=1}^n p_j\left\|x_j\right\|^{p-2}\left[v,x_j\right]\geq \sum_{j=1}^n p_j\left\|x_j\right\|^p\right\}.$$

If $(X,\langle\cdot,\cdot\rangle)$ is an inner product space then $Sla\left(\left\|\cdot\right\|^p,\mathbf{x},\mathbf{p}\right)$ can be described by

$$Sla\left(\left\|\cdot\right\|^p,\mathbf{x},\mathbf{p}\right)=\left\{v\in X \mid \left\langle v,\sum_{j=1}^n p_j\left\|x_j\right\|^{p-2}x_j\right\rangle\geq \sum_{j=1}^n p_j\left\|x_j\right\|^p\right\}$$

and if the family $\{x_j\}_{j=1,...,n}$ is orthogonal, then obviously, by the Pythagoras theorem, we have that the sum $\sum_{j=1}^n x_j$ belongs to $Sla\left(\left\|\cdot\right\|,\mathbf{x}\right)$ and therefore to $Sla\left(\left\|\cdot\right\|^p,\mathbf{x},\mathbf{p}\right)$ for any $p\geq 1$ and any probability distribution $\mathbf{p}=(p_1,...,p_n)\in\mathbb{P}^n$.

We can state now the following results that provides a generalization of Slater's inequality as well as a counterpart for it.

THEOREM 2.2 (Dragomir, 2012 [**58**]). *Let $f:X\to\mathbb{R}$ be a convex function on the real linear space X, $\mathbf{x}=(x_1,...,x_n)\in X^n$ an n-tuple of vectors and $\mathbf{p}=(p_1,...,p_n)\in\mathbb{P}^n$ a probability distribution. Then for any $v\in Sla_+(f,\mathbf{x},\mathbf{p})$ we have the inequalities*

$$\nabla_-f(v)(v)-\sum_{i=1}^n p_i\nabla_-f(v)(x_i)\geq f(v)-\sum_{i=1}^n p_if(x_i)\geq 0. \tag{2.9}$$

PROOF. If we write the gradient inequality for $v\in Sla_+(f,\mathbf{x},\mathbf{p})$ and x_i, then we have that

$$\nabla_-f(v)(v-x_i)\geq f(v)-f(x_i)\geq\nabla_+f(x_i)(v-x_i) \tag{2.10}$$

for any $i\in\{1,...,n\}$.

By multiplying (2.10) with $p_i\geq 0$ and summing over i from 1 to n we get

$$\sum_{i=1}^n p_i\nabla_-f(v)(v-x_i)\geq f(v)-\sum_{i=1}^n p_if(x_i)\geq\sum_{i=1}^n p_i\nabla_+f(x_i)(v-x_i). \tag{2.11}$$

Now, since $v\in Sla_+(f,\mathbf{x},\mathbf{p})$, then the right hand side of (2.11) is nonnegative, which proves the second inequality in (2.9).

By the superadditivity of the Gâteaux derivative $\nabla_-f(\cdot)(\cdot)$ in the second variable we have

$$\nabla_-f(v)(v)-\nabla_-f(v)(x_i)\geq\nabla_-f(v)(v-x_i),$$

which, by multiplying with $p_i\geq 0$ and summing over i from 1 to n, produces the inequality

$$\nabla_-f(v)(v)-\sum_{i=1}^n p_i\nabla_-f(v)(x_i)\geq\sum_{i=1}^n p_i\nabla_-f(v)(v-x_i). \tag{2.12}$$

Utilising (2.11) and (2.12) we deduce the desired result (2.9). ∎

REMARK 2.1. The above result has the following form for normed linear spaces. Let $(X,\|\cdot\|)$ be a normed linear space, $\mathbf{x}=(x_1,...,x_n)\in X^n$ an n-tuple of vectors from X and $\mathbf{p}=(p_1,...,p_n)\in\mathbb{P}^n$ a probability distribution. Then for any vector $v\in X$ with the property

$$\sum_{j=1}^n p_j\left\|x_j\right\|^{p-2}\langle v,x_j\rangle_s\geq\sum_{j=1}^n p_j\left\|x_j\right\|^p,\ p\geq 1, \tag{2.13}$$

we have the inequalities

$$p\left[\|v\|^p-\sum_{j=1}^n p_j\|x_j\|^{p-2}\langle v,x_j\rangle_i\right]\geq\|v\|^p-\sum_{j=1}^n p_j\|x_j\|^p\geq 0. \tag{2.14}$$

Rearranging the first inequality in (2.14) we also have that

$$(p-1)\|v\|^p+\sum_{j=1}^n p_j\|x_j\|^p\geq p\sum_{j=1}^n p_j\|x_j\|^{p-2}\langle v,x_j\rangle_i\,. \tag{2.15}$$

If the space is smooth, then the condition (2.13) becomes

$$\sum_{j=1}^n p_j\|x_j\|^{p-2}[v,x_j]\geq\sum_{j=1}^n p_j\|x_j\|^p\,,\ p\geq 1, \tag{2.16}$$

implying the inequality

$$p\left[\|v\|^p-\sum_{j=1}^n p_j\|x_j\|^{p-2}[v,x_j]\right]\geq\|v\|^p-\sum_{j=1}^n p_j\|x_j\|^p\geq 0. \tag{2.17}$$

Notice also that the first inequality in (2.17) is equivalent with

$$(p-1)\|v\|^p+\sum_{j=1}^n p_j\|x_j\|^p\geq p\sum_{j=1}^n p_j\|x_j\|^{p-2}[v,x_j]\qquad\left(\geq p\sum_{j=1}^n p_j\|x_j\|^p\geq 0\right). \tag{2.18}$$

The following corollary is of interest:

COROLLARY 2.3 (Dragomir, 2012 [**58**]). *Let $f:X\to\mathbb{R}$ be a convex function on the real linear space X, $\mathbf{x}=(x_1,...,x_n)\in X^n$ an n-tuple of vectors and $\mathbf{p}=(p_1,...,p_n)\in\mathbb{P}^n$ a probability distribution. If*

$$\sum_{i=1}^n p_i\nabla_+f(x_i)(x_i)\geq(<)\,0 \tag{2.19}$$

and there exists a vector $s\in X$ with

$$\sum_{i=1}^n p_i\nabla_{+(-)}f(x_i)(s)\geq(\leq)\,1 \tag{2.20}$$

then

$$\begin{aligned}\nabla_-f\left(\sum_{j=1}^n p_j\nabla_+f(x_j)(x_j)\,s\right)&\left(\sum_{j=1}^n p_j\nabla_+f(x_j)(x_j)\,s\right)\\&-\sum_{i=1}^n p_i\nabla_-f\left(\sum_{j=1}^n p_j\nabla_+f(x_j)(x_j)\,s\right)(x_i)\\&\geq f\left(\sum_{j=1}^n p_j\nabla_+f(x_j)(x_j)\,s\right)-\sum_{i=1}^n p_if(x_i)\geq 0.\end{aligned} \tag{2.21}$$

PROOF. Assume that $\sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \geq 0$ and $\sum_{i=1}^{n} p_i \nabla_+ f(x_i)(s) \geq 1$ and define $v := \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\, s$. We claim that $v \in Sla_+ (f, \mathbf{x}, \mathbf{p})$.

By the subadditivity and positive homogeneity of the mapping $\nabla_+ f(\cdot)(\cdot)$ in the second variable we have

$$\begin{aligned}
&\sum_{i=1}^{n} p_i \nabla_+ f(x_i)(v - x_i) \\
\geq\ & \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(v) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \sum_{i=1}^{n} p_i \nabla_+ f(x_i) \left(\sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\, s \right) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j) \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(s) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j) \left[\sum_{i=1}^{n} p_i \nabla_+ f(x_i)(s) - 1 \right] \geq 0,
\end{aligned}$$

as claimed. Applying Theorem 2.2 for this v we get the desired result.

If $\sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) < 0$ and $\sum_{i=1}^{n} p_i \nabla_- f(x_i)(s) \leq 1$ then for

$$w := \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\, s$$

we also have that

$$\begin{aligned}
&\sum_{i=1}^{n} p_i \nabla_+ f(x_i)(w - x_i) \\
\geq\ & \sum_{i=1}^{n} p_i \nabla_+ f(x_i) \left(\sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j)\, s \right) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \sum_{i=1}^{n} p_i \nabla_+ f(x_i) \left(\left(- \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j) \right)(-s) \right) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \left(- \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j) \right) \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(-s) - \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(x_i) \\
=\ & \left(- \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j) \right) \left(1 + \sum_{i=1}^{n} p_i \nabla_+ f(x_i)(-s) \right) \\
=\ & \left(- \sum_{j=1}^{n} p_j \nabla_+ f(x_j)(x_j) \right) \left(1 - \sum_{i=1}^{n} p_i \nabla_- f(x_i)(s) \right) \geq 0
\end{aligned}$$

where, for the last equality we have used the property (2.3). Therefore $w \in Sla_+ (f, \mathbf{x}, \mathbf{p})$ and by Theorem 2.2 we get the desired result. ∎

It is natural to consider the case of normed spaces.

REMARK 2.2. Let $(X, \|\cdot\|)$ be a normed linear space, $\mathbf{x} = (x_1, ..., x_n) \in X^n$ an n-tuple of vectors from X and $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ a probability distribution. Then for any vector $s \in X$ with the property that

$$p \sum_{i=1}^{n} p_i \|x_i\|^{p-2} \langle s, x_i \rangle_s \geq 1, \tag{2.22}$$

we have the inequalities

$$p^p \|s\|^{p-1} \left(\sum_{j=1}^{n} p_j \|x_j\|^p \right)^{p-1} \left(p \|s\| \sum_{j=1}^{n} p_j \|x_j\|^p - \sum_{j=1}^{n} p_j \langle x_j, s \rangle_i \right) \geq p^p \|s\|^p \left(\sum_{j=1}^{n} p_j \|x_j\|^p \right)^p - \sum_{j=1}^{n} p_j \|x_j\|^p \geq 0.$$

The case of smooth spaces can be easily derived from the above, however the details are left to the interested reader.

3. THE CASE OF FINITE DIMENSIONAL LINEAR SPACES

Consider now the finite dimensional linear space $X = \mathbb{R}^m$ and assume that C is an open convex subset of $\mathbb{R}^m$. Assume also that the function $f : C \to \mathbb{R}$ is differentiable and convex on C. Obviously, if $x = (x^1, ..., x^m) \in C$ then for any $y = (y^1, ..., y^m) \in \mathbb{R}^m$ we have

$$\nabla f(x)(y) = \sum_{k=1}^{m} \frac{\partial f(x^1, ..., x^m)}{\partial x^k} \cdot y^k$$

For the convex function $f : C \to \mathbb{R}$ and a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in C^n$ with $x_i = (x_i^1, ..., x_i^m)$ with $i \in \{1, ..., n\}$, we consider the sets

$$Sla(f, \mathbf{x}, C) := \left\{ v \in C \mid \sum_{k=1}^{m} \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot v^k \geq \sum_{k=1}^{m} \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot x_i^k \text{ for all } i \in \{1, ..., n\} \right\} \tag{3.1}$$

and

$$Sla(f, \mathbf{x}, \mathbf{p}, C) := \left\{ v \in C \mid \sum_{i=1}^{n} \sum_{k=1}^{m} p_i \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot v^k \geq \sum_{i=1}^{n} \sum_{k=1}^{m} p_i \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot x_i^k \right\} \tag{3.2}$$

where $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ is a given probability distribution.

As in the previous section the sets $Sla(f, \mathbf{x}, C)$ and $Sla(f, \mathbf{x}, \mathbf{p}, C)$ are convex and closed subsets of $\text{clo}(C)$, the closure of C. Also $\{x_1, ..., x_n\} \subset Sla(f, \mathbf{x}, C) \subset Sla(f, \mathbf{x}, \mathbf{p}, C)$ for any $\mathbf{p} = (p_1, ..., p_n) \in \mathbb{P}^n$ a probability distribution.

PROPOSITION 3.1. *Let $f : C \to \mathbb{R}$ be a convex function on the open convex set C in the finite dimensional linear space $\mathbb{R}^m$, $(x_1, ..., x_n) \in C^n$ an n-tuple of vectors and $(p_1, ..., p_n) \in$*

$\mathbb{P}^n$ *a probability distribution. Then for any* $v = (v^1, ..., v^n) \in Sla(f, \mathbf{x}, \mathbf{p}, C)$ *we have the inequalities*

$$\sum_{k=1}^{m} \frac{\partial f(v^1, ..., v^m)}{\partial x^k} \cdot v^k - \sum_{i=1}^{n} \sum_{k=1}^{m} p_i \frac{\partial f(x_i^1, ..., x_i^m)}{\partial x^k} \cdot v^k \geq f\left(v^1, ..., v^n\right) - \sum_{i=1}^{n} p_i f\left(x_i^1, ..., x_i^m\right) \geq 0. \tag{3.3}$$

The unidimensional case, i.e., $m = 1$ is of interest for applications. We will state this case with the general assumption that $f : I \to \mathbb{R}$ is a convex function on an *open* interval I. For a given n-tuple of vectors $\mathbf{x} = (x_1, ..., x_n) \in I^n$ we have

$$Sla_{+(-)}(f, \mathbf{x}, I) := \left\{v \in I \mid f'_{+(-)}(x_i) \cdot (v - x_i) \geq 0 \text{ for all } i \in \{1, ..., n\}\right\}$$

and

$$Sla_{+(-)}(f, \mathbf{x}, \mathbf{p}, \mathbf{I}) := \left\{v \in I \middle| \sum_{i=1}^{n} p_i f'_{+(-)}(x_i) \cdot (v - x_i) \geq 0 \right\},$$

where $(p_1, ..., p_n) \in \mathbb{P}^n$ is a probability distribution. These sets inherit the general properties pointed out in Lemma 2.1. Moreover, if we make the assumption that $\sum_{i=1}^{n} p_i f'_+(x_i) \neq 0$ then for $\sum_{i=1}^{n} p_i f'_+(x_i) > 0$ we have

$$Sla_+(f, \mathbf{x}, \mathbf{p}, \mathbf{I}) = \left\{v \in I \middle| v \geq \frac{\sum_{i=1}^{n} p_i f'_+(x_i) x_i}{\sum_{i=1}^{n} p_i f'_+(x_i)} \right\}$$

while for $\sum_{i=1}^{n} p_i f'_+(x_i) < 0$ we have

$$v = \left\{v \in I \middle| v \leq \frac{\sum_{i=1}^{n} p_i f'_+(x_i) x_i}{\sum_{i=1}^{n} p_i f'_+(x_i)} \right\}.$$

Also, if we assume that $f'_+(x_i) \geq 0$ for all $i \in \{1, ..., n\}$ and $\sum_{i=1}^{n} p_i f'_+(x_i) > 0$ then

$$v_s := \frac{\sum_{i=1}^{n} p_i f'_+(x_i) x_i}{\sum_{i=1}^{n} p_i f'_+(x_i)} \in I$$

due to the fact that $x_i \in I$ and I is a convex set.

PROPOSITION 3.2. *Let* $f : I \to \mathbb{R}$ *be a convex function on an open interval* I. *For a given* n*-tuple of vectors* $\mathbf{x} = (x_1, ..., x_n) \in I^n$ *and* $(p_1, ..., p_n) \in \mathbb{P}^n$ *a probability distribution we have*

$$f'_-(v) \left(v - \sum_{i=1}^{n} p_i x_i\right) \geq f(v) - \sum_{i=1}^{n} p_i f(x_i) \geq 0 \tag{3.4}$$

for any $v \in Sla_+(f, \mathbf{x}, \mathbf{p}, \mathbf{I})$.

In particular, if we assume that $\sum_{i=1}^{n} p_i f'_+(x_i) \neq 0$ *and*

$$\frac{\sum_{i=1}^{n} p_i f'_+(x_i) x_i}{\sum_{i=1}^{n} p_i f'_+(x_i)} \in I$$

then

$$f'_{-}\left(\frac{\sum_{i=1}^{n} p_i f'_{+}(x_i)\, x_i}{\sum_{i=1}^{n} p_i f'_{+}(x_i)}\right)\left[\frac{\sum_{i=1}^{n} p_i f'_{+}(x_i)\, x_i}{\sum_{i=1}^{n} p_i f'_{+}(x_i)} - \sum_{i=1}^{n} p_i x_i\right] \geq f\left(\frac{\sum_{i=1}^{n} p_i f'_{+}(x_i)\, x_i}{\sum_{i=1}^{n} p_i f'_{+}(x_i)}\right) - \sum_{i=1}^{n} p_i f(x_i) \geq 0 \tag{3.5}$$

Moreover, if $f'_{+}(x_i) \geq 0$ *for all* $i \in \{1, ..., n\}$ *and* $\sum_{i=1}^{n} p_i f'_{+}(x_i) > 0$ *then (3.5) holds true as well.*

REMARK 3.1. We remark that the first inequality in (3.5) provides a reverse inequality for the classical result due to Slater.

4. SOME APPLICATIONS FOR f-DIVERGENCES

It is obvious that the above definition of $I_f(\mathbf{p}, \mathbf{q})$ can be extended to any function $f : [0, \infty) \to \mathbb{R}$ however the positivity condition will not generally hold for normalized functions and $\mathbf{p}, \mathbf{q} \in R_{+}^{n}$ *with* $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i$.

For a normalized convex function $f : [0, \infty) \to \mathbb{R}$ and two probability distributions $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ we define the set

$$Sla_{+}(f, \mathbf{p}, \mathbf{q}) := \left\{ v \in [0, \infty) \middle| \sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right) \cdot \left(v - \frac{p_i}{q_i}\right) \geq 0 \right\}. \tag{4.1}$$

Now, observe that

$$\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right) \cdot \left(v - \frac{p_i}{q_i}\right) \geq 0$$

is equivalent with

$$v \sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right) \geq \sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right). \tag{4.2}$$

If $\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right) > 0$, then (4.2) is equivalent with

$$v \geq \frac{\sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)}$$

therefore in this case

$$Sla_{+}(f, \mathbf{p}, \mathbf{q}) = \begin{cases} [0, \infty) & \text{if } \sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right) < 0 \\ \left[\frac{\sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)}, \infty\right) & \text{if } \sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right) \geq 0. \end{cases} \tag{4.3}$$

If $\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right) < 0$, then (4.2) is equivalent with

$$v \leq \frac{\sum_{i=1}^{n} p_i f'_{+}\left(\frac{p_i}{q_i}\right)}{\sum_{i=1}^{n} q_i f'_{+}\left(\frac{p_i}{q_i}\right)}$$

therefore

$$Sla_{+}(f,\mathbf{p},\mathbf{q})=\begin{cases}\left[0,\frac{\sum_{i=1}^{n}p_{i}f'_{+}\left(\frac{p_{i}}{q_{i}}\right)}{\sum_{i=1}^{n}q_{i}f'_{+}\left(\frac{p_{i}}{q_{i}}\right)}\right] & \text{if } \sum_{i=1}^{n}p_{i}f'_{+}\left(\frac{p_{i}}{q_{i}}\right)\leq 0\\ \emptyset & \text{if } \sum_{i=1}^{n}p_{i}f'_{+}\left(\frac{p_{i}}{q_{i}}\right)>0.\end{cases} \tag{4.4}$$

Utilising the extended f-divergences notation, we can state the following result:

THEOREM 4.1 (Dragomir, 2012 [**58**]). *Let $f:[0,\infty)\to\mathbb{R}$ be a normalized convex function and $\mathbf{p},\mathbf{q}\in\mathbb{P}^{n}$ two probability distributions. If $v\in Sla_{+}(f,\mathbf{p},\mathbf{q})$ then we have*

$$f'_{-}(v)(v-1)\geq f(v)-I_{f}(\mathbf{p},\mathbf{q})\geq 0. \tag{4.5}$$

In particular, if we assume that $I_{f'_{+}}(\mathbf{p},\mathbf{q})\neq 0$ and

$$\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}\in[0,\infty)$$

then

$$f'_{-}\left(\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}\right)\left[\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}-1\right]\geq f\left(\frac{I_{f'_{+}(\cdot)(\cdot)}(\mathbf{p},\mathbf{q})}{I_{f'_{+}}(\mathbf{p},\mathbf{q})}\right)-I_{f}(\mathbf{p},\mathbf{q})\geq 0. \tag{4.6}$$

Moreover, if $f'_{+}\left(\frac{p_{i}}{q_{i}}\right)\geq 0$ for all $i\in\{1,...,n\}$ and $I_{f'_{+}}(\mathbf{p},\mathbf{q})>0$ then (4.6) holds true as well.

The proof follows immediately from Proposition 3.2 and the details are omitted.

The K. Pearson χ^{2}*-divergence* is obtained for the convex function $f(t)=(1-t)^{2}$, $t\in\mathbb{R}$ and given by

$$\chi^{2}(\mathbf{p},\mathbf{q}):=\sum_{j=1}^{n}q_{j}\left(\frac{p_{j}}{q_{j}}-1\right)^{2}=\sum_{j=1}^{n}\frac{(p_{j}-q_{j})^{2}}{q_{j}}=\sum_{j=1}^{n}\frac{p_{i}^{2}}{q_{i}}-1. \tag{4.7}$$

The *Kullback-Leibler divergence* can be obtained for the convex function $f:(0,\infty)\to\mathbb{R}$, $f(t)=t\ln t$ and is defined by

$$KL(\mathbf{p},\mathbf{q}):=\sum_{j=1}^{n}q_{j}\cdot\frac{p_{j}}{q_{j}}\ln\left(\frac{p_{j}}{q_{j}}\right)=\sum_{j=1}^{n}p_{j}\ln\left(\frac{p_{j}}{q_{j}}\right). \tag{4.8}$$

If we consider the convex function $f:(0,\infty)\to\mathbb{R}$, $f(t)=-\ln t$, then we observe that

$$I_{f}(\mathbf{p},\mathbf{q}):=\sum_{i=1}^{n}q_{i}f\left(\frac{p_{i}}{q_{i}}\right)=-\sum_{i=1}^{n}q_{i}\ln\left(\frac{p_{i}}{q_{i}}\right)=\sum_{i=1}^{n}q_{i}\ln\left(\frac{q_{i}}{p_{i}}\right)=KL(\mathbf{q},\mathbf{p}). \tag{4.9}$$

For the function $f(t) = -\ln t$ we have obviously have that

$$Sla(-\ln, \mathbf{p}, \mathbf{q}) := \left\{ v \in [0, \infty) | -\sum_{i=1}^{n} q_i \left(\frac{p_i}{q_i}\right)^{-1} \cdot \left(v - \frac{p_i}{q_i}\right) \geq 0 \right\} = \left\{ v \in [0, \infty) | \, v \sum_{i=1}^{n} \frac{q_i^2}{p_i} - 1 \leq 0 \right\} = \left[0, \frac{1}{\chi^2(\mathbf{q}, \mathbf{p}) + 1}\right]. \tag{4.10}$$

Utilising the first part of the Theorem 4.1 we can state the following

PROPOSITION 4.2. *Let $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ two probability distributions. If $v \in \left[0, \frac{1}{\chi^2(\mathbf{q},\mathbf{p})+1}\right]$ then we have*

$$\frac{1-v}{v} \geq -\ln(v) - KL(\mathbf{q}, \mathbf{p}) \geq 0. \tag{4.11}$$

In particular, for $v = \frac{1}{\chi^2(\mathbf{q},\mathbf{p})+1}$ we get

$$\chi^2(\mathbf{q}, \mathbf{p}) \geq \ln\left[\chi^2(\mathbf{q}, \mathbf{p}) + 1\right] - KL(\mathbf{q}, \mathbf{p}) \geq 0. \tag{4.12}$$

If we consider now the function $f : (0, \infty) \to \mathbb{R}$, $f(t) = t \ln t$, then $f'(t) = \ln t + 1$ and

$$Sla((\cdot)\ln(\cdot), \mathbf{p}, \mathbf{q}) := \left\{ v \in [0, \infty) | \sum_{i=1}^{n} q_i \left(\ln\left(\frac{p_i}{q_i}\right) + 1\right) \cdot \left(v - \frac{p_i}{q_i}\right) \geq 0 \right\} = \left\{ v \in [0, \infty) | \, v \sum_{i=1}^{n} q_i \left(\ln\left(\frac{p_i}{q_i}\right) + 1\right) - \sum_{i=1}^{n} p_i \cdot \left(\ln\left(\frac{p_i}{q_i}\right) + 1\right) \geq 0 \right\} = \{ v \in [0, \infty) | \, v(1 - KL(\mathbf{q}, \mathbf{p})) \geq 1 + KL(\mathbf{p}, \mathbf{q}) \}. \tag{4.13}$$

We observe that if $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ two probability distributions such that $0 < KL(\mathbf{q}, \mathbf{p}) < 1$, then

$$Sla((\cdot)\ln(\cdot), \mathbf{p}, \mathbf{q}) = \left[\frac{1 + KL(\mathbf{p}, \mathbf{q})}{1 - KL(\mathbf{q}, \mathbf{p})}, \infty\right).$$

If $KL(\mathbf{q}, \mathbf{p}) \geq 1$ then $Sla((\cdot)\ln(\cdot), \mathbf{p}, \mathbf{q}) = \emptyset$.

By the use of Theorem 4.1 we can state now the following

PROPOSITION 4.3. *Let $\mathbf{p}, \mathbf{q} \in \mathbb{P}^n$ two probability distributions such that $0 < KL(\mathbf{q}, \mathbf{p}) < 1$. If $v \in \left[\frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})}, \infty\right)$ then we have*

$$(\ln v + 1)(v - 1) \geq v \ln(v) - KL(\mathbf{p}, \mathbf{q}) \geq 0. \tag{4.14}$$

In particular, for $v = \frac{1+KL(\mathbf{p},\mathbf{q})}{1-KL(\mathbf{q},\mathbf{p})}$ we get

$$\left(\ln\left[\frac{1 + KL(\mathbf{p}, \mathbf{q})}{1 - KL(\mathbf{q}, \mathbf{p})}\right] + 1\right)\left(\frac{1 + KL(\mathbf{p}, \mathbf{q})}{1 - KL(\mathbf{q}, \mathbf{p})} - 1\right) \geq \frac{1 + KL(\mathbf{p}, \mathbf{q})}{1 - KL(\mathbf{q}, \mathbf{p})} \ln\left[\frac{1 + KL(\mathbf{p}, \mathbf{q})}{1 - KL(\mathbf{q}, \mathbf{p})}\right] - KL(\mathbf{p}, \mathbf{q}) \geq 0. \tag{4.15}$$

Similar results can be obtained for other divergence measures of interest such as the *Jeffreys divergence, Hellinger discrimination,* etc...However the details are left to the interested reader.

CHAPTER 12

Approximation of f-Divergence Via Ostrowski Type Inequalities

1. Some Preliminary Results

The difference between two probability measures p, q on a set $A = \{\alpha_i | 1 \leq i \leq n\}$ is commonly measured in a variety of ways. Denote by p_i, q_i the associated point probabilities for the event $\alpha_i \in A$. To avoid triviality we assume that $p_i + q_i > 0$ for each i. The *variational distance* (ℓ_1–distance) and *information divergence* (Kullback–Leibler divergence) between the distributions p and q are defined respectively by

$$V(p,q) := \sum_{i=1}^{n} |p_i - q_i|,$$

$$D(p,q) := \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i}.$$

Another measure, which proves a useful benchmark in our analysis, is the chi–squared divergence of p, q, which is defined by

$$D_{\chi^2}(p,q) := \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1 = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}.$$

The last two measures are unfortunately infinite if $p_i > 0$ but $q_i = 0$ for some i. This complication is obviated in the *triangular discrimination* between p and q, which is defined as in [**127**] by

$$\Delta(p,q) := \sum_{i=1}^{n} \frac{|p_i - q_i|^2}{p_i + q_i}.$$

A generalization of this measure, parameterized by a natural number υ, is

$$\Delta_\upsilon(p,q) := \sum_{i=1}^{n} \frac{|p_i - q_i|^{2\upsilon}}{(p_i + q_i)^{2\upsilon - 1}},$$

which we refer to as *triangular discrimination of order* υ (see [**127**]). Another common choice is the *Hellinger discrimination*

$$h^2(p,q) := \frac{1}{2} \sum_{i=1}^{n} \left(\sqrt{p_i} - \sqrt{q_i}\right)^2.$$

For applications it is important to know how these divergences compare with one another. The basic relations between V, Δ and h^2 are

$$\frac{1}{2} V^2(p,q) \leq \Delta(p,q) \leq V(p,q)$$

and

$$2h^2(p,q) \leq \Delta(p,q) \leq 4h^2(p,q)$$

(see LeCam [**99**] and Dacunha–Castelle [**27**]). From these we may deduce that

$$\frac{1}{8}V^2(p,q) \le h^2(p,q) \le \frac{1}{2}V(p,q).$$

The coefficients in these inequalities are best possible (*cf.* [**127**]).

The first half of this result has been improved by Kraft [**92**], who showed that

$$\frac{1}{8}V^2(p,q) \le h^2(p,q)\left(1 - \frac{1}{2}h^2(p,q)\right).$$

We note also the important inequality

$$D(p,q) \ge -2\ln\left(1 - h^2(p,q)\right)$$

(see Dacunha–Castelle [**27**]). It follows from this that

$$D(p,q) \ge 2h^2(p,q).$$

Again the coefficient 2 is best–possible (see [**127**]).

The key to unity in this diversity is that all the discrepancy measures considered above are particular instances of Csiszár f–divergences. If $f : [0,\infty) \to \mathbf{R}$ is convex, the *Csiszár f-divergence* between p and q is defined by

$$I_f(p,q) := \sum_{i=1}^{n} q_i f\left(p_i/q_i\right) \tag{1.1}$$

(see Csiszár [**22**]–[**24**]). Thus the family $(f_s)_{s\ge 1}$ of functions with

$$f_s(u) = |u-1|^s (u+1)^{1-s}$$

gives rise to variational distance when $s = 1$, triangular discrimination when $s = 2$ and triangular discrimination of order v when $s = 2v$ (see [**127**]). The choice $f(u) = \frac{1}{2}(\sqrt{u}-1)^2$ gives rise to Hellinger discrimination and $f(u) = u\ln u$ to Kullback-Leibler divergence. The chi–squared divergence is given by $f(u) = (u-1)^2$.

For all of the above choices $f(1) = 0$, so that $I_f(p,p) = 0$. The convexity of f then ensures that $I_f(p,q)$ is nonnegative.

In Section 2 we derive, by the use of Ostrowski's integral inequality for absolutely continuous mappings with essentially bounded first derivative, an approximation for the Csiszár f–divergence in terms of an integral mean. With many concrete examples this provides very simple approximations. Section 3 considers some of the examples noted above and Section 4 the case when each pair p_i, q_i are very close. Finally, in Section 5, we look at applications to mutual information.

It needs to be stressed that as these estimates lose most of the detailed information involved in the values p_i, q_i, the approximations, while very simple, can also be very crude.

2. An Inequality for Csiszár f–Divergence

In some applications it is convenient to make use of definition (1.1) for functions $f : [0,\infty) \to \mathbf{R}$ which are continuous but not necessarily convex. An illustrative example is given in Section 3. Accordingly our main result, Theorem 1 below, does not assume convexity.

We assume in what follows that there exist real numbers r, R with

$$0 < r \le p_i/q_i \le R < \infty$$

for all $i \in \{1,...,n\}$. Note that if $r > 1$, then $p_i > q_i$ for each i, which gives $1 = \sum_i p_i > \sum_i q_i = 1$, a contradiction. Hence $r \le 1$. A similar argument gives $R \ge 1$.

Further suppose that the restriction of f to the compact interval $[r,R]$ is absolutely continuous. We derive an approximation to the Csiszár f–divergence in terms of the integral mean of f over $[r,R]$. We shall show in Theorem 1 below that if p and q are close in the sense that $R-r$ is small, then the integral mean

$$\frac{1}{R-r}\int_r^R f(t)dt$$

approximates the Csiszár f–divergence to first order.

We make use of Ostrowski's integral inequality, which states the following. See [**77**] for a short proof and some applications to numerical integration and special means.

THEOREM 2.1. *Assume that* $g:[a,b]\to R$ *is absolutely continuous with* $g^{'}\in L_\infty[a,b]$*, that is, that* $\left\|g^{'}\right\|_\infty := ess\sup_{t\in[a,b]}\left|g^{'}(t)\right| < \infty.$

Then

$$\left|g(x)-\frac{1}{b-a}\int_a^b g(t)dt\right| \le \left[\frac{1}{4}+\left(\frac{x-\frac{a+b}{2}}{b-a}\right)^2\right](b-a)\left\|g^{'}\right\|_\infty$$

for all $x\in[a,b]$.

A further key result is due to Diaz and Metcalf (see [**109**, p. 61]).

THEOREM 2.2. *Suppose* $a_k(\neq 0)$ *and* b_k $(k=1,...,n)$ *are real numbers satisfying* $m\le b_k/a_k\le M$*. Then*

$$\sum_{k=1}^n b_k^2 + mM\sum_{k=1}^n a_k^2 \le (M+m)\sum_{k=1}^n a_k b_k.$$

Equality holds if and only if for each k *either* $b_k = ma_k$ *or* $b_k = Ma_k$.

We shall make use of a slight extension of this.

Suppose the conditions of the Diaz–Metcalf result hold and $t_k>0$ for $k=1,...,n$. Then

$$\sum_{k=1}^n t_k b_k^2 + mM\sum_{k=1}^n t_k a_k^2 \le (M+m)\sum_{k=1}^n t_k a_k b_k.$$

Equality holds if and only if for each k either $b_k = ma_k$ or $b_k = Ma_k$.

We have for $k=1,2,\ldots,n$ that

$$(b_k/a_k - m)(M - b_k/a_k)t_k a_k^2 \ge 0.$$

The desired result follows on summation over k.

THEOREM 2.3 (Dragomir et al., 2000 [**69**]). *Assume that* $f:[r,R]\to\mathbf{R}$ *is absolutely continuous on* $[r,R]$ *and* $f^{'}\in L_\infty[r,R]$*. Then*

$$\left|I_f(p,q)-\frac{1}{R-r}\int_r^R f(t)dt\right| \tag{2.1}$$
$$\le\left[\frac{1}{4}+\frac{1}{(R-r)^2}\left\{D_{\chi^2}(p,q)+\left(\frac{R+r}{2}-1\right)^2\right\}\right](R-r)\left\|f^{'}\right\|_\infty$$
$$\le\frac{1}{2}(R-r)\left\|f^{'}\right\|_\infty.$$

PROOF. By Ostrowski's integral inequality, we have

$$\left|f\left(\frac{p_i}{q_i}\right)-\frac{1}{R-r}\int_r^R f(t)dt\right|\leq\left[\frac{1}{4}+\left(\frac{\frac{p_i}{q_i}-\frac{R+r}{2}}{R-r}\right)^2\right](R-r)\left\|f'\right\|_\infty$$

for each $i\in\{1,...,n\}$.

We may multiply by q_i, sum the resultant inequalities and use the generalized triangle inequality to obtain

$$\begin{aligned}&\left|I_f(p,q)-\frac{1}{R-r}\int_r^R f(t)dt\right|\\&\leq\sum_{i=1}^n q_i\left|f\left(\frac{p_i}{q_i}\right)-\frac{1}{R-r}\int_r^R f(t)dt\right|\\&\leq\left[\frac{1}{4}+\frac{1}{(R-r)^2}\sum_{i=1}^n q_i\left(\frac{p_i}{q_i}-\frac{R+r}{2}\right)^2\right](R-r)\left\|f'\right\|_\infty.\end{aligned}$$

Since

$$\begin{aligned}\sum_{i=1}^n q_i\left(\frac{p_i}{q_i}-\frac{R+r}{2}\right)^2&=\sum_{i=1}^n\frac{p_i^2}{q_i}-(R+r)+\left(\frac{R+r}{2}\right)^2\\&\leq\sum_{i=1}^n\frac{p_i^2}{q_i}-1+\left(\frac{R+r}{2}-1\right)^2\\&=D_{\chi^2}(p,q)+\left(\frac{R+r}{2}-1\right)^2,\end{aligned}$$

this yields the first inequality in (2.1).

For the second, set $b_k=\sqrt{p_k/q_k}$ and $a_k=\sqrt{q_k/p_k}$ $(k=1,...,n)$. Then $a_k/b_k=p_k/q_k\in[r,R]$ $(k\in\{1,...,n\})$. On applying Proposition 1 for $t_k=p_k$ $(k=1,...,n)$, we get

$$\sum_{k=1}^n p_k\left(\sqrt{\frac{p_k}{q_k}}\right)^2+rR\sum_{k=1}^n p_k\left(\sqrt{\frac{q_k}{p_k}}\right)^2\leq(r+R)\sum_{k=1}^n p_k\sqrt{\frac{p_k}{q_k}}\cdot\sqrt{\frac{q_k}{p_k}},$$

or equivalently

$$\sum_{k=1}^n\frac{p_k^2}{q_k}+rR\leq R+r.$$

Thus

$$D_{\chi^2}(p,q)\leq r+R-rR-1=(1-r)(R-1)$$

and so

$$\frac{1}{4}+\frac{1}{(R-r)^2}\left[D_{\chi^2}(p,q)+\frac{1}{4}(R+r-2)^2\right]\leq\frac{1}{2}$$

and the theorem is proved. ∎

COROLLARY 2.4 (Dragomir et al., 2000 [**69**]). *Let f satisfy the conditions of Theorem 2.3. If $\varepsilon>0$ and*

$$0\leq R-r\leq 2\epsilon/\left\|f'\right\|_\infty,$$

then

$$\left| I_f(p,q) - \frac{1}{R-r}\int_r^R f(t)dt \right| \le \epsilon.$$

Theorem 1 can be reformulated to emphasize the approximation aspect.

Let $f : [0,2] \to \mathbf{R}$ be absolutely continuous with $f^{'} \in L_\infty[0,2]$. If $\eta \in (0,1)$ and $p(\eta)$, $q(\eta)$ are such that

$$\left| \frac{p_i(\eta)}{q_i(\eta)} - 1 \right| \le \eta$$

for all $i \in \{1,...,n\}$, then

$$I_f(p(\eta),q(\eta)) = \frac{1}{2\eta}\int_{1-\eta}^{1+\eta} f(t)dt + R_f(p,q,\eta)$$

and the remainder $R_f(p,q,\eta)$ satisfies

$$|R_f(p,q,\eta)| \le \frac{\eta}{2}\left[1 + \frac{1}{\eta^2}D_{\chi^2}\left(p(\eta),q(\eta)\right)\right]\left\|f^{'}\right\|_\infty \le \eta\left\|f^{'}\right\|_\infty.$$

This follows by Theorem 2.3 with the choices $R = 1+\eta$ and $r = 1-\eta$ ($\eta \in (0,1)$).

3. Particular cases

For Kullback–Leibler distance, we take $f(u) = u \ln u$. With this choice we have $\left\|f^{'}\right\|_\infty = \ln(eR)$ and

$$\begin{aligned}
\int_r^R f(t)dt &= \frac{1}{4}\left[R^2 \ln R^2 - r^2 \ln r^2 - \left(R^2 - r^2\right)\right] \\
&= \frac{R^2 - r^2}{4}\ln\left[\left(\frac{(R^2)^{(R^2)}}{(r^2)^{(r^2)}}\right)^{1/(R^2-r^2)} \cdot \frac{1}{e}\right] \\
&= \frac{R^2 - r^2}{4} I\left[\left(R^2, r^2\right)\right],
\end{aligned}$$

where the identric mean $I(a,b)$ for positive arguments is given by

$$I(a,b) := \begin{cases} a & \text{if } b = a \\ \frac{1}{e}\left(\frac{b^b}{a^a}\right)^{1/(b-a)} & \text{if } b \ne a. \end{cases}$$

The conclusion of Theorem 1 reads

$$\begin{aligned}
&\left| D(p,q) - \frac{R+r}{4}\ln\left[I\left(R^2,r^2\right)\right]\right| \\
&\le \left[\frac{1}{4} + \frac{1}{(R-r)^2}\left[D_{\chi^2}(p,q) + \left(\frac{R+r}{2} - 1\right)^2\right]\right](R-r)\ln(eR) \\
&\le \frac{1}{2}(R-r)\ln(eR).
\end{aligned} \tag{3.1}$$

If we take the concave map $f:(0,\infty)\to\mathbf{R}$ given by $f(u)=\ln u$, then we have

$$I_f(p,q)=\sum_{k=1}^{n} q_i \ln\frac{p_i}{q_i}=-D(q,p).$$

With this choice $\left\|f^{'}\right\|_{\infty}=1/r$ and the identric mean reappears through

$$\frac{1}{R-r}\int_r^R f(t)dt=\ln\left[I(r,R)\right].$$

Theorem 1 provides

$$\begin{aligned}
&\left|D(q,p)-\ln\left[\frac{1}{I(r,R)}\right]\right| \\
&\leq\left[\frac{1}{4}+\frac{1}{(R-r)^2}\left[D_{\chi^2}(p,q)+\left(\frac{R+r}{2}-1\right)^2\right]\right]\left(\frac{R}{r}-1\right) \\
&\leq\frac{1}{2}\left(\frac{R}{r}-1\right).
\end{aligned}\tag{3.2}$$

For Hellinger discrimination $f(u)=\left(\sqrt{u}-1\right)^2/2$, so

$$f^{'}(u)=\frac{\sqrt{u}-1}{2\sqrt{u}},\quad f^{''}(u)=\frac{1}{4u}$$

for $u\in(0,\infty)$ and

$$\left\|f^{'}\right\|_{\infty}=\sup_{u\in[r,R]}\left|f^{'}(u)\right|=\left|f^{'}(R)\right|=\frac{\sqrt{R}-1}{2\sqrt{R}}.$$

Also

$$\frac{1}{R-r}\int_r^R f(t)dt=\frac{R+r}{4}-\frac{2}{3}\cdot\frac{R+\sqrt{rR}+r}{\sqrt{r}+\sqrt{R}}+\frac{1}{2},$$

and inequality (2.1) becomes

$$\begin{aligned}
&\left|h^2(p,q)-\left[\frac{R+r}{4}-\frac{2}{3}\cdot\frac{R+\sqrt{rR}+r}{\sqrt{r}+\sqrt{R}}+\frac{1}{2}\right]\right| \\
&\leq\left[\frac{1}{4}+\frac{1}{(R-r)^2}\left\{D_{\chi^2}(p,q)+\left(\frac{R+r}{2}-1\right)^2\right\}\right](R-r)\left[\frac{\sqrt{R}-1}{2\sqrt{R}}\right] \\
&\leq\frac{1}{4\sqrt{R}}(R-r)(\sqrt{R}-1).
\end{aligned}$$

For variational distance, $f(u)=|u-1|$, which is absolutely continuous on $[r,R]$. We have

$$f^{'}(u):=\begin{cases}-1 & \text{if } u\in(r,1)\\ 1 & \text{if } u\in(1,R),\end{cases}$$

so that

$$\left\|f^{'}\right\|_{\infty}=\sup_{t\in[r,R]}\left|f^{'}(t)\right|=1.$$

Further

$$\frac{1}{R-r}\int_r^R f(t)dt = \frac{1}{R-r}\left[\int_r^1 (1-u)du + \int_1^R (u-1)du\right]$$
$$= \frac{1}{R-r}\left[\frac{(r-1)^2}{2} + \frac{(R-1)^2}{2}\right]$$
$$= \frac{1}{R-r}\left[\frac{(R-r)^2}{4} + \left(\frac{r+R}{2}-1\right)^2\right].$$

Theorem 1 provides

$$\left|V(p,q) - \frac{1}{R-r}\left[\frac{(R-r)^2}{4} + \left(\frac{r+R}{2}-1\right)^2\right]\right|$$
$$\leq \left[\frac{1}{4} + \frac{1}{(R-r)^2}\left\{D_{\chi^2}(p,q) + \left(\frac{R+r}{2}-1\right)^2\right\}\right](R-r)$$
$$\leq \frac{1}{2}(R-r).$$

Our final example relates to triangular discrimination, which arises with $f(u) = (u-1)^2/(u+1)$. We have

$$f(u) = u + 1 - \frac{4u}{u+1}, \quad f'(u) = 1 - \frac{4}{(u+1)^2}$$

for $u \in [0, \infty)$, so that

$$\left\|f'\right\|_\infty = \sup_{u\in[r,R]} |f'(u)| = |f'(R)| = \frac{(R-1)(R+1)}{(R+1)^2}.$$

Also

$$\frac{1}{R-r}\int_r^R f(u)du = \frac{R+r}{2} + \ln\left(\frac{R+1}{r+1}\right)^{4/(R-r)} - 3$$

and Theorem 1 provides

$$\left|\Delta(p,q) - \left[\frac{R+r}{2} + \ln\left(\frac{R+1}{r+1}\right)^{4/(R-r)} - 3\right]\right|$$
$$\leq \left[\frac{1}{4} + \frac{1}{(R-r)^2}\left\{D_{\chi^2}(p,q) + \left(\frac{R+r}{2}-1\right)^2\right\}\right]\frac{(R-r)(R-1)(R+3)}{(R+1)^2}$$
$$\leq \frac{1}{2}\frac{(R-r)(R-1)(R+3)}{(R+1)^2}.$$

4. SOME NUMERICAL EXAMPLES

One situation of practical interest is where p_i and q_i are close, so that we have $p_i = p_i(\varepsilon)$, $q_i = q_i(\varepsilon)$ and

$$\left|\frac{p_i(\varepsilon)}{q_i(\varepsilon)} - 1\right| \leq \varepsilon \; \varepsilon \in (0,1) \tag{4.1}$$

for all $i \in \{1, ..., n\}$. With $R = \varepsilon + 1$ and $r = 1 - \varepsilon$, we obtain from (3.1) that

$$\begin{aligned}\left| D(p(\varepsilon), q(\varepsilon)) - \frac{1}{2}\ln\left[I\left((1+\varepsilon)^2, (1-\varepsilon)^2\right)\right]\right| &\le \frac{\varepsilon}{2}\left[1 + \frac{1}{\varepsilon^2}D_{\chi^2}(p(\varepsilon), q(\varepsilon))\right]\ln\left[e(1+\varepsilon)\right] \\ &\le \varepsilon \ln\left[e(1+\varepsilon)\right].\end{aligned}$$

Consequently if $p(\varepsilon)$, $q(\varepsilon)$ are in the sense of (4.1), we can approximate the Kullback-Leibler distance $D(p(\varepsilon), q(\varepsilon))$ by $(1/2)\ln\left[I\left((1+\varepsilon)^2, (1-\varepsilon)^2\right)\right]$ and the error of the approximation is less than

$$E(\varepsilon) := \varepsilon \ln[e(1+\varepsilon)].$$

From (3.2), we derive

$$\begin{aligned}\left| D(q(\varepsilon), p(\varepsilon)) - \ln\left[\frac{1}{I(1-\varepsilon, 1+\varepsilon)}\right]\right| &\le \frac{1}{2}\frac{\varepsilon}{1-\varepsilon}\left[1 + \frac{1}{\varepsilon^2}D_{\chi^2}(p(\varepsilon), q(\varepsilon))\right] \\ &\le \frac{\varepsilon}{1-\varepsilon}\end{aligned}$$

for $\varepsilon \in (0, 1)$.

Consequently for $p(\varepsilon), q(\varepsilon)$ satisfying (4.1), we can approximate the Kullback–Leibler distance $D(p(\varepsilon), q(\varepsilon))$ by $\ln\left[I^{-1}(1-\varepsilon, 1+\varepsilon)\right]$ and the error of the approximation is less than $\varepsilon/(1-\varepsilon)$ for $\varepsilon \in (0, 1)$.

5. APPLICATION TO MUTUAL INFORMATION

We consider mutual information, which is a measure of the amount of information that one random variable provides about another. It is the reduction of uncertainty about one variable due to knowledge of the other (see, for example, [**21**]).

Consider two discrete–valued random variables X and Y with a joint probability mass function $t(x, y)$ and marginal probability mass functions $p(x)$ ($x \in \mathcal{X}$) and $q(y)$ ($y \in \mathcal{Y}$). The mutual information is the relative entropy between the joint distribution and the product distribution, that is,

$$I(X; Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} t(x, y)\ln\left[\frac{t(x, y)}{p(x)q(y)}\right] = D\left(t(x, y), p(x)q(y)\right),$$

where as before $D(\cdot, \cdot)$ denotes Kullback–Leibler distance.

We assume that

$$s \le \frac{t(x, y)}{p(x)q(y)} \le S \quad \text{for all } (x, y) \in \mathcal{X} \times \mathcal{Y}. \tag{5.1}$$

Much as with r, R we have $s \le 1 \le S$.

We also may consider mutual information in a chi–squared sense, that is,

$$I_{\chi^2}(X; Y) := \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{t^2(x, y)}{p(x)q(y)} - 1.$$

Inequality (3.1) yields the following proposition.

If t, p and q satisfy (5.1), then

$$\begin{aligned}
&\left| I(X;Y) - \frac{s+S}{4} \ln\left[I\left(s^2, S^2\right)\right]\right| \\
&\leq \left[\frac{1}{4} + \frac{1}{(S-s)^2}\left[I_{\chi^2}(X;Y) + \left(\frac{s+S}{2} - 1\right)^2\right]\right](S-s)\ln[eS] \\
&\leq \frac{1}{2}(S-s)\ln(eS).
\end{aligned}$$

The condition $t(x,y) = p(x)p(y)$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$ means that the random variables X and Y are independent. We may refer to them as "quasi–independent" if

$$\left|\frac{t(x,y)}{p(x)q(y)} - 1\right| \leq \delta \qquad (\delta \in (0,1))$$

for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$. When this occurs, we can approximate the mutual information $I(X;Y)$ by

$$\frac{1}{2}\left[\frac{(1+\delta)^2 \ln(1+\delta)^2 - (1-\delta)^2 \ln(1-\delta)^2}{4\delta} - 1\right] \quad (\delta \in (0,1))$$

with an error less than $E(\delta)$ for $t \in (0,1)$.

6. PRELIMINARIES

The Csiszár f–divergence between $p = (p_1, \ldots, p_n)$ and $q = (q_1, \ldots, q_n)$ is defined by the functional

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

Two important instances, which we shall invoke shortly, are the variational distance $V(p,q)$ and the chi–squared divergence $D_{\chi^2}(p,q)$, for which

$$f(u) = |u-1|^m$$

with $m = 1, 2$ respectively. We address the situation in which there exist constants r, R with

(6.1) $$0 < r < 1 < R < \infty \text{ and } r \leq p_i/q_i \leq R \text{ for } i = 1, \ldots, n.$$

LEMMA 6.1. *If (6.1) is satisfied, then*

$$V(p,q) \leq \frac{2(R-1)(1-r)}{R-r} \leq \frac{R-r}{2}.$$

The first inequality is an equality if and only if for each i either $p_i/q_i = r$ or $p_i/q_i = R$. The second inequality is an equality if and only if $R + r = 2$.

We start with the following proposition which provides an Ostrowski–type inequality for mappings of bounded variation. This has been established by the author in the preprint [**34**]. We give a simple proof.

PROPOSITION 6.2 (Dragomir, 1999 [**34**]). *Suppose $g : [a,b] \to \mathbf{R}$ is of bounded variation on $[r,R]$. Then for all $x \in [a,b]$,*

(6.2) $$\left|\int_a^b g(t)dt - g(x)(b-a)\right| \leq \left[\frac{b-a}{2} + \left|x - \frac{a+b}{2}\right|\right] \bigvee_a^b (g) \leq (b-a) \bigvee_a^b (g),$$

where $\bigvee_a^b (g)$ denotes the total variation of g on $[a,b]$. The constant $1/2$ is best–possible.

PROOF. Using the integration by parts formula for a Riemann–Stieltjes integral, we have that $\int_a^x (t-a)dg(t)$ and $\int_x^b (t-b)dg(t)$ exist and that

$$\int_a^x (t-a)dg(t) = g(x)(x-a) - \int_a^x g(t)dt$$

and

$$\int_x^b (t-b)dg(t) = g(x)(b-x) - \int_x^b g(t)dt$$

for all $x \in [a,b]$.

Addition provides

$$g(x)(b-a) - \int_a^b g(t)dt = \int_a^x (t-a)dg(t) + \int_x^b (t-b)dg(t)$$

for all $x \in [a,b]$.

Now if $p, g : [a,b] \to \mathbf{R}$ with p continuous and g of bounded variation, then $\int_a^b p(t)dg(t)$ exists and

$$\left| \int_a^b p(t)dg(t) \right| = \sup_{t \in [a,b]} |p(t)| \bigvee_a^b (g).$$

Hence

$$\begin{aligned}
\left| g(x)(b-a) - \int_a^b g(t)dt \right| &= \left| \int_a^x (t-a)dg(t) + \int_x^b (t-b)dg(t) \right| \\
&\le \left| \int_a^x (t-a)dg(t) \right| + \left| \int_x^b (t-b)dg(t) \right| \\
&\le \sup_{t \in [a,x]} |t-a| \bigvee_a^x (g) + \sup_{t \in [x,b]} |t-b| \bigvee_x^b (g) \\
&= (x-a) \bigvee_a^x (g) + (b-x) \bigvee_x^b (g) \\
&\le \max\{x-a, b-x\} \left[\bigvee_a^x (g) + \bigvee_x^b (g) \right] \\
&= \left[\frac{b-a}{2} + \left| x - \frac{a+b}{2} \right| \right] \bigvee_a^b (g),
\end{aligned}$$

and the first inequality in (6.2) is proved. The second follows, since $|x-(a+b)/2| \le (b-a)/2$.

Suppose that (6.2) holds with a constant $c > 0$, that is,

$$\left| \int_a^b g(t)dt - g(x)(b-a) \right| \le \left[c(b-a) + \left| x - \frac{a+b}{2} \right| \right] \bigvee_a^b (g) \tag{6.3}$$

for all $x \in [a, b]$, and define $g_1 : [a, b] \to \mathbf{R}$ by

$$g_1(x) = \begin{cases} 1 & \text{if } x = (a+b)/2 \\ 0 & \text{otherwise.} \end{cases}$$

Then g_1 is of bounded variation on $[a, b]$ and

$$\bigvee_a^b (g_1) = 2, \quad \int_a^b g_1(t)dt = 0.$$

Put $g = g_1$ and $x = (a + b)/2$ in (6.3). Then we get $1 \le 2c$, which shows that $c = 1/2$ is best–possible. ∎

PROPOSITION 6.3. *If $g : [a, b] \to \mathbf{R}$ is of bounded variation, then for all $x_1, x_2 \in [a, b]$,*

$$\left| \int_a^b g(t)dt - \frac{b-a}{2} \sum_{i=1}^{2} g(x_i) \right| \le \left[\frac{b-a}{2} + \frac{1}{2} \sum_{i=1}^{2} \left| x_i - \frac{a+b}{2} \right| \right] \bigvee_a^b (g). \tag{6.4}$$

PROOF. This follows by putting $x = x_i$ in Proposition 6.2, summing over i and then using the triangle inequality. ∎

LEMMA 6.4. *Suppose $f : [r, R] \to \mathbf{R}$ is differentiable, so that f' is of bounded variation on $[r, R]$. If $r < 1 < R$ and $x \in [r, R]$, then*

$$\left| f(x) - f(1) - \frac{x-1}{2} \left[f'(1) + f'(x) \right] \right| \le |x - 1| \bigvee_r^R (f'). \tag{6.5}$$

PROOF. For $x \ge 1$, we set $g = f'$, $x_1 = a = 1$ and $x_2 = b = x$ in (6.4) to derive

$$\left| f(x) - f(1) - \frac{x-1}{2} \left[f'(1) + f'(x) \right] \right| \le (x - 1) \bigvee_1^x (f') \le (x - 1) \bigvee_r^R (f').$$

Similarly if $x < 1$, we set $g = f'$, $x_1 = a = x$ and $x_2 = b = 1$ in (6.4) to derive

$$\left| f(1) - f(x) - \frac{1-x}{2} \left[f'(1) + f'(x) \right] \right| \le (1 - x) \bigvee_x^1 (f') \le (1 - x) \bigvee_r^R (f').$$

The desired result follows in both cases. ∎

7. BOUNDS IN TERMS OF TOTAL VARIATION

THEOREM 7.1 (Dragomir et al., 2001 [**68**]). *If $f:[r,R]\to\mathbf{R}$ is of bounded variation and (6.1) applies, then*

$$\left|I_f(p,q)-\frac{1}{R-r}\int_r^R f(t)dt\right| \tag{7.1}$$
$$\le\left[\frac{1}{2}+\frac{1}{R-r}\sum_{k=1}^n\left|p_i-\frac{r+R}{2}\cdot q_i\right|\right]\bigvee_r^R(f)$$
$$\le\left\{\frac{1}{2}+\frac{1}{R-r}\left[V(p,q)+\left|\frac{r+R}{2}-1\right|\right]\right\}\bigvee_r^R(f)$$
$$\le\left[1+\frac{1}{R-r}\cdot V(p,q)\right]\bigvee_r^R(f)$$
$$\le\frac{3}{2}\bigvee_r^R(f).$$

PROOF. The choices $g=f$, $x=p_i/q_i$ $(i=1,...,n)$, $a=r$, $b=R$ in (6.2) give

$$\left|f\left(\frac{p_i}{q_i}\right)-\frac{1}{R-r}\int_r^R f(t)dt\right|\le\left[\frac{1}{2}+\frac{1}{R-r}\left|\frac{p_i}{q_i}-\frac{r+R}{2}\right|\right]\bigvee_r^R(f)$$

for all $i\in\{1,...,n\}$.

If we multiply by q_i and sum over i, we obtain *via* the generalized triangle inequality that

$$\left|I_f(p,q)-\frac{1}{R-r}\int_r^R f(t)dt\right|\le\left[\frac{1}{2}+\frac{1}{R-r}\sum_{i=1}^n q_i\left|\frac{p_i}{q_i}-\frac{r+R}{2}\right|\right]\bigvee_r^R(f),$$

whence we have the first inequality in (7.1).

The second follows from

$$\sum_{i=1}^n q_i\left|\frac{p_i}{q_i}-1-\left(\frac{r+R}{2}-1\right)\right|\le\sum_{i=1}^n q_i\left|\frac{p_i}{q_i}-1\right|+\left|\frac{r+R}{2}-1\right|$$
$$=V(p,q)+\left|\frac{r+R}{2}-1\right|$$

and the third from

$$\left|\frac{r+R}{2}-1\right|\le\frac{R-r}{2}.$$

The final inequality follows by Lemma 6.1. ∎

The following corollary emphasizes better the approximation aspect of the theorem.

COROLLARY 7.2. *Let $f:[0,2]\to\mathbf{R}$ be a mapping of bounded variation. If $\eta\in(0,1)$ and $p(\eta)$ and $q(\eta)$ are probability distributions satisfying*

$$\left|\frac{p_i(\eta)}{q_i(\eta)}-1\right|\le\eta\ \text{ for all } i\in\{1,...,n\},$$

then

$$I_f(p(\eta), q(\eta) = \frac{1}{2\varepsilon} \int_{1-\eta}^{1+\eta} f(t)dt + R_f(p, q, \eta)$$

and the reminder term R_f satisfies

$$|R_f(p, q, \eta)| \le \frac{1}{2}\left[1 + \frac{1}{\eta} V(p(\eta), q(\eta))\right] \bigvee_{1-\eta}^{1+\eta}(f).$$

We note that the best inequality we can get from (6.2) is

$$\left| \int_a^b g(t)dt - g\left(\frac{a+b}{2}\right)(b-a) \right| \le \frac{b-a}{2} \bigvee_a^b (g), \tag{7.2}$$

which arises for $x = (a+b)/2$.

Suppose

$$f^*(u) := f(1) + (u-1)f'\left(\frac{1+u}{2}\right).$$

Similarly define

$$f^\dagger(u) := f(1) + \frac{u-1}{2} f'(u).$$

THEOREM 7.3 (Dragomir et al., 2001 [**68**]). *Suppose $f : [r, R] \to \mathbf{R}$ is differentiable and so f' is of bounded variation. If (6.1) applies, then*

$$|I_f(p, q) - I_{f^*}(p, q)| \le \frac{1}{2} V(p, q) \bigvee_r^R (f') \le \frac{R-1}{4} \bigvee_r^R (f'), \tag{7.3}$$

$$\left|I_f(p, q) - I_{f^\dagger}(p, q)\right| \le V(p, q) \bigvee_r^R (f') \le \frac{R-1}{2} \bigvee_r^R (f'). \tag{7.4}$$

PROOF. Taking (7.2) with $g = f'$, $a = 1$, and $b = x \in [r, R]$ gives

$$|f(x) - f^*(x)| \le \frac{|x-1|}{2} \bigvee_1^x (f') \le \frac{|x-1|}{2} \bigvee_r^R (f')$$

for all $x \in [r, R]$. The first inequality in (7.3) follows by putting $x = p_i/q_i$, multiplying by q_i, summing over $i = 1, \ldots, n$ and using the generalized triangle inequality. The second inequality is given by Lemma 6.1.

The proof of (7.4) follows similarly from (6.5). ∎

Both parts may be viewed in terms of approximation. Thus for (7.3) we have the following.

COROLLARY 7.4. *Suppose $f : [0, 2] \to \mathbf{R}$ has its first derivative of bounded variation. If $\eta \in (0, 1)$ and $p(\eta)$, $q(\eta)$ are probability distributions satisfying,*

$$\left|\frac{p_i(\eta)}{q_i(\eta)} - 1\right| \le \eta \text{ for all } i \in \{1, ..., n\},$$

then

$$I_f(p(\eta), q(\eta) = I_{f^*}((p(\eta), q(\eta)) + R_f(p, q, \eta)$$

and the reminder R_f is such that

$$|R_f(p,q,\eta)| \le \frac{1}{2} V(p(\eta),q(\eta)) \bigvee_{1-\eta}^{1+\eta}(f') \le \frac{\eta}{2} \bigvee_{1-\eta}^{1+\eta}(f').$$

8. EXAMPLES

Suppose (6.1) holds and $f : [r,R] \to \mathbf{R}$ is given by $f(u) = u \ln u$, so that $I_f(p,q)$ is the Kullback–Leibler distance

$$D(p,q) := \sum_{i=1}^{n} p_i \ln(p_i/q_i).$$

We have

$$\int_r^R f(t)dt = \frac{1}{4}\left[R^2 \ln R^2 - r^2 \ln r^2 - \left(R^2 - r^2\right)\right]$$
$$= \frac{R^2 - r^2}{4} \ln\left[I\left(R^2, r^2\right)\right],$$

where $I(a,b)$ is the identric mean of two positive numbers a, b and is given by

$$I(a,b) := \begin{cases} a & \text{if } b = a \\ \frac{1}{e}\left(\frac{b^b}{a^a}\right)^{1/(b-a)} & \text{otherwise} . \end{cases}$$

Also

$$\bigvee_r^R(f) = \int_a^b \left|f'(t)\right| dt = \int_a^b |\ln(et)|\, dt =: \lambda(r,R).$$

If $0 < r \le 1/e$, then

$$\lambda(r,R) = \int_r^{1/e} [-\ln(et)]\, dt + \int_{1/e}^R \ln(et)dt$$
$$= -\frac{1}{e}\int_r^{1/e} \ln(et)d(et) + \frac{1}{e}\int_{1/e}^R \ln(et)d(et).$$

Since

$$\int_\alpha^\beta \ln x\, dx = \ln I(\alpha,\beta) \ \text{ for } \alpha,\beta > 0,$$

we have

$$\lambda(r,R) = -\frac{1}{e}\ln\left[I\left(r,e^{-1}\right)\right] + \frac{1}{e}\ln\left[I(e^{-1},R)\right] = \ln\left[\frac{I(e^{-1},R)}{I(r,e^{-1})}\right]^{1/e}.$$

If on the other hand $1/e < r < 1$, then

$$\lambda(r,R) = \int_r^R \ln(et)dt = \frac{1}{e}\ln I(r,R) = \ln\left[I(r,R)\right]^{1/e}.$$

Thus (7.1) gives

$$\begin{aligned}\left|D(p,q)-\frac{R+r}{4}\ln\left[I\left(R^2,r^2\right)\right]\right| &\leq \left[\frac{1}{2}+\frac{1}{R-r}\sum_{i=1}^{n}\left|p_i-\frac{r+R}{2}\cdot q_i\right|\right]\lambda(r,R)\\ &\leq \left\{\frac{1}{2}+\frac{1}{R-r}\left[V(p,q)+\left|\frac{r+R}{2}-1\right|\right]\right\}\lambda(r,R)\\ &\leq \left[1+\frac{1}{R-r}V(p,q)\right]\lambda(r,R)\\ &\leq \frac{3}{2}\lambda(r,R).\end{aligned}$$

Also

$$I_{f^*}(p,q)=\sum_{i=1}^{n}(p_i-q_i)\ln\left(\frac{p_i+q_i}{2q_i}\right)$$

and

$$\bigvee_r^R(f')=\int_r^R\left|f''(t)\right|dt=\int_r^R\frac{dt}{t}=\ln\left(\frac{R}{r}\right).$$

Hence by (7.3) we have

$$\begin{aligned}\left|D(p,q)-\sum_{i=1}^{n}(p_i-q_i)\ln\left(\frac{p_i+q_i}{2q_i}\right)\right| &\leq \frac{1}{2}V(p,q)\ln\left(\frac{R}{r}\right)\\ &\leq \frac{R-r}{4}(\ln R-\ln r)\\ &= \frac{(R-r)^2}{4L(r,R)},\end{aligned}$$

where $L(a,b)$ is the logarithmic mean which for positive arguments a, b is given by

$$L(a,b):=\begin{cases} a & \text{if } b=a\\ \frac{b-a}{\ln b-\ln a} & \text{otherwise.}\end{cases}$$

Therefore we have the inequality

$$\left|D(p,q)-\sum_{i=1}^{n}(p_i-q_i)\ln\left(\frac{p_i+q_i}{2q_i}\right)\right|\leq\frac{1}{4}\cdot\frac{R-r}{L(r,R)}V(p,q)\leq\frac{(R-r)^2}{4L(r,R)}.$$

Finally suppose $f:[r,R]\to\mathbf{R}$ is given by $f(u)=|u-1|$, so $I_f(p,q)$ becomes the variational distance

$$V(p,q):=\sum_{i=1}^{n}|p_i-q_i|.$$

We have

$$\begin{aligned}
\frac{1}{R-r}\int_r^R f(t)dt &= \frac{1}{R-r}\int_r^R |u-1|\,du \\
&= \frac{1}{R-r}\left[\int_r^1 (1-u)du + \frac{1}{R-r}\int_1^R (u-1)\,du\right] \\
&= \frac{1}{R-r}\left[\frac{(r-1)^2}{2} + \frac{(R-1)^2}{2}\right] \\
&= \frac{1}{R-r}\left[\frac{(R-r)^2}{4} + \left(\frac{r+R}{2}-1\right)^2\right]
\end{aligned}$$

and

$$\bigvee_r^R(f) = \bigvee_r^1(f) + \bigvee_1^R(f) = 1-r+R-1 = R-r.$$

By (7.1) we have

$$\begin{aligned}
&\left|V(p,q) - \frac{1}{R-r}\left[\frac{(R-r)^2}{4} + \left(\frac{r+R}{2}-1\right)^2\right]\right| \\
&\le \left[\frac{1}{2} + \frac{1}{R-r}\sum_{i=1}^n \left|p_i - \frac{r+R}{2}\cdot q_i\right|\right](R-r) \\
&\le \left\{\frac{1}{2} + \frac{1}{R-r}\left[V(p,q) + \left|\frac{r+R}{2}-1\right|\right]\right\}(R-r) \\
&\le \left[1 + \frac{1}{R-r}V(p,q)\right](R-r) \\
&\le \frac{3(R-r)}{2}.
\end{aligned}$$

CHAPTER 13

Approximation of f-Divergence Via Midpoint and Trapezoid Inequalities

1. Introduction

By $A = \{a_i | 1 \leq i \leq n\}$ we denote an n-set, the alphabet. Distributions over A, always assumed to be probability distributions are typically denoted by p, q and the associated point probabilities are denoted by p_i, q_i.

Variational distance (l_1-distance) and *information divergence* (Kullback-Leibler divergence) are defined as usual (see for example [**21**])

$$V(p,q) = \sum_{i=1}^{n} |p_i - q_i|, \tag{1.1}$$

$$D(p,q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}. \tag{1.2}$$

Here, log denotes natural logarithm.

As in [**127**], we define *triangular discrimination* between p and q by

$$\Delta(p,q) = \sum_{i=1}^{n} \frac{|p_i - q_i|^2}{p_i + q_i}. \tag{1.3}$$

A variant of this measure, depending on a natural number υ as a parameter, are called *triangular discrimination of order* υ [**127**], will also be considered.

It is defined by the equation

$$\Delta_\upsilon(p,q) = \sum_{i=1}^{n} \frac{|p_i - q_i|^{2\upsilon}}{(p_i + q_i)^{2\upsilon - 1}}. \tag{1.4}$$

For $\upsilon = 1$, we are back to (1.3), i.e. $\Delta_1 = \Delta$.

All measures of discrepancy considered above are particular instances of Csiszár f-divergences. Recall, cf. Csiszár [**22**], [**23**], [**24**] that for a convex function $f : [0, \infty) \to R$ the *Csiszár f-divergence* between p and q is defined by

$$I_f(p,q) := \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right). \tag{1.5}$$

The family of function $(f_s)_{s \geq 1}$ with $f_s(u) = |u-1|^s (u+1)^{1-s}$ gives rise to variational distance $V(s = 1)$, triangular discrimination $\Delta(s = 2)$ and triangular discrimination of order υ, Δ_υ $(s = 2\upsilon)$ [**127**].

Among the most popular choices we mention $f(u) = \frac{1}{2}(\sqrt{u}-1)^2$ which gives rise to the *Hellinger discrimination* h^2:

$$h^2(p,q) = \frac{1}{2}\sum_{i=1}^{n}\left(\sqrt{p_i}-\sqrt{q_i}\right)^2. \tag{1.6}$$

For $f(u) = u\log u$, we obtain also the Kullback-Leibler divergence $D(p,q)$.

The basic relations between V, Δ and h^2 are the following (see also [**127**])

$$\frac{1}{2}V^2(p,q) \le \Delta(p,q) \le V(p,q) \tag{1.7}$$

(cf. (11) in [**127**]),

$$2h^2(p,q) \le \Delta(p,q) \le 4h^2(p,q) \tag{1.8}$$

(cf. LeCam [**99**] and Dacunha-Castelle [**27**]), and, lastly, the relation

$$\frac{1}{8}V^2(p,q) \le h^2(p,q) \le \frac{1}{2}V(p,q) \tag{1.9}$$

(follows from the two first).

The occurring coefficient are best possible (cf. [**127**]).

Kraft [**92**] improved part of this by pointing out that

$$\frac{1}{8}V^2(p,q) \le h^2(p,q)\left(1-\frac{1}{2}h^2(p,q)\right). \tag{1.10}$$

In Dacunha-Castelle [**27**] we find the important inequality

$$D(p,q) \ge -2\log\left(1-h^2(p,q)\right), \tag{1.11}$$

in particular

$$D(p,q) \ge 2h^2(p,q). \tag{1.12}$$

This is the best possible as the inequality $D(p,q) \ge ch^2$ cannot hold for any $c < 2$ ([**127**]).

In the recent paper [**66**], by the use of Ostrowski's integral inequality for absolutely continuous functions, we proved the following approximation result:

$$I_f(p,q) = \frac{1}{2\varepsilon}\int_{1-\varepsilon}^{1+\varepsilon} f(t)dt + R_f(p,q,\varepsilon), \quad \varepsilon \in (0,1) \tag{1.13}$$

where the remainder $R_f(p,q,\varepsilon)$ satisfies the estimate

$$\begin{aligned} R_f(p,q,\varepsilon) &\le \frac{1}{2}\left[1+\frac{1}{\varepsilon^2}D_{\chi^2}(p,q)\right]\varepsilon\left\|f'\right\|_\infty \\ &\le \varepsilon\left\|f'\right\|_\infty \end{aligned} \tag{1.14}$$

and provided

(i) f is absolutely continuous on $[0,2]$ whose derivative $f' \in L_\infty[0,2]$ and $\left\|f'\right\|_\infty := ess\sup_{t\in[0,2]}\left|f'(t)\right|$;

(ii) p and q satisfy the condition

$$\left|\frac{p_i}{q_i}-1\right| \le \varepsilon \text{ for all } i \in \{1,...,n\}. \tag{1.15}$$

Application for particular divergence measures (variational distance, information divergence, triangular discrimination, Hellinger discrimination etc..) were also given.

In the present chapter we point out other approximations for the Csiszár f-divergence of order $O(\varepsilon^2)$ and $O(\varepsilon^3)$ by the use of some mid-point inequalities. Application for particular divergence measure are also given.

2. An Inequality for Csiszár f-Divergence

Assume that $f : [0, \infty) \to R$ is a continuous mapping (not necessarily convex) and consider the Csiszár f-divergence between p and q as considered above

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

We assume in what follows that there exists the real numbers r, R so that $0 < r \le r_i := \frac{p_i}{q_i} \le R < \infty$ for all $i \in \{1, ..., n\}$. Obviously $1 \le r \le R$.

Using this basically assumption and supposing that f restricted to the compact interval $[r, R]$ has the first derivative (the second derivative) absolutely continuous and $f''(f''') \in L_\infty[r, R]$, then we are able to point out two approximation results of the Csiszár f-divergence as follows.

If p, q are close in the following sense:

$$\left|\frac{p_i}{q_i} - 1\right| \le \eta,\ \eta > 0 \quad (\eta\text{-small})$$

then we have

$$I_f(p, q) \approx f(1) + \sum_{i=1}^{n} (p_i - q_i) f'\left(\frac{p_i + q_i}{2q_i}\right),$$

and the accuracy is of order $O(\eta^2)$ respectively $O(\eta^3)$.

To be more precise, we state the first result.

THEOREM 2.1 (Dragomir et al., 2001 [**65**]). *Assume that the mapping $f : [0, \infty) \to R$ is so that $f' : [r, R] \to R$ is absolutely continuous on $[r, R]$ and $f'' \in L_\infty[r, R]$, i.e., $\left\|f'\right\|_\infty := ess \sup_{t\in[r,R]} \left|f'(t)\right| < \infty$. Then we have the inequality:*

$$\begin{aligned} &\left|I_f(p, q) - f(1) - \sum_{i=1}^{n} (p_i - q_i) f'\left(\frac{p_i + q_i}{2q_i}\right)\right| \\ &\le \frac{1}{4}\left\|f''\right\|_\infty D_{\chi^2}(p, q) \le \frac{1}{4}\left\|f''\right\|_\infty (R - 1)(1 - r) \\ &\le \frac{1}{16}\left\|f''\right\|_\infty (R - r)^2, \end{aligned} \tag{2.1}$$

where

$$D_{\chi^2}(p, q) := \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1 = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i}$$

is the Chi-Square divergence of p, q.

PROOF. Firstly, let us recall Ostrowski's integral inequality for absolutely continuous functions $g : [a, b] \to R$ whose derivatives $g' \in L_\infty[a, b]$

$$\left|g(x) - \frac{1}{b - a}\int_a^b g(t)dt\right| \le \left[\frac{1}{4} + \left(\frac{x - \frac{a+b}{2}}{b - a}\right)^2\right](b - a)\left\|g'\right\|_\infty, \tag{2.2}$$

for all $x \in [a, b]$.

For a simple proof of this, see [**77**] where some applications in numerical integration and for special means are given.

It is obvious that the best inequality you can have in (2.2) is for $x = \frac{a+b}{2}$ obtaining the mid-point inequality

$$\left| g\left(\frac{a+b}{2}\right)(b-a) - \int_a^b g(t)dt \right| \leq \frac{1}{4}(b-a)^2 \left\| g' \right\|_\infty . \tag{2.3}$$

Choose in (2.3) $g = f'$, $a = 1$, $b = x \in [r, R]$ (note that $r \leq 1 \leq R$) to get

$$\left| f(x) - f(1) - (x-1)f'\left(\frac{1+x}{2}\right) \right| \leq \frac{1}{4}(x-1)^2 \left\| f'' \right\|_\infty . \tag{2.4}$$

for all $x \in [r, R]$.

If we put $x = \frac{p_i}{q_i} \in [r, R]$ in (2.4), then we obtain

$$\left| q_i f\left(\frac{p_i}{q_i}\right) - q_i f(1) - (p_i - q_i) f'\left(\frac{p_i + q_i}{2q_i}\right) \right| \leq \frac{1}{4}\left(\frac{p_i - q_i}{q_i}\right)^2 q_i \left\| f'' \right\|_\infty \tag{2.5}$$

for all $i \in \{1, ..., n\}$.

Summing (2.5), over i from 1 to n and using the generalized triangle inequality, we get

$$\left| I_f(p,q) - f(1) - \sum_{i=1}^n (p_i - q_i) f'\left(\frac{p_i + q_i}{2q_i}\right) \right| \leq \frac{1}{4}\left\| f'' \right\|_\infty \sum_{i=1}^n \frac{(p_i - q_i)^2}{q_i} \tag{2.6}$$

and the first inequality in (2.1) is proved.

We prove now that (see also [**66**])

$$D_{\chi^2}(p,q) \leq (R-1)(1-r).$$

For this purpose, we use the following well known result due to Diaz and Metcalf (see for example [**109**, p. 61]):

Let $t_k > 0$ $(k = 1, ..., n)$ with $\sum_{k=1}^n t_k = 1$. If $a_k (\neq 0)$ and b_k $(k = 1, ..., n)$ are real numbers and if

$$m \leq \frac{b_k}{a_k} \leq M \tag{2.7}$$

for $k = 1, ..., n$; then

$$\sum_{k=1}^n t_k b_k^2 + mM \sum_{k=1}^n t_k a_k^2 \leq (M+m) \sum_{k=1}^n t_k a_k b_k . \tag{2.8}$$

Equality holds in (2.8) if and only if for each k, $1 \leq k \leq n$ either $b_k = m a_k$ or $b_k = M a_k$.

Define

$$b_k = \sqrt{\frac{p_k}{q_k}}, a_k = \sqrt{\frac{q_k}{p_k}}, k = 1, ..., n.$$

Then

$$\frac{a_k}{b_k} = \frac{p_k}{q_k} = r_k \in [r, R]$$

for all $k \in \{1, ..., n\}$. Applying the inequality (2.8) for $t_k = p_k \ (k = 1, ..., n)$, we get

$$\sum_{k=1}^{n} p_k \left(\sqrt{\frac{p_k}{q_k}}\right)^2 + rR \sum_{k=1}^{n} p_k \left(\sqrt{\frac{q_k}{p_k}}\right)^2 \le (r+R) \sum_{k=1}^{n} p_k \sqrt{\frac{p_k}{q_k}} \cdot \sqrt{\frac{q_k}{p_k}}$$

which is equivalent to

$$\sum_{k=1}^{n} \frac{p_k^2}{q_k} + rR \le r + R$$

and the inequality (2.6) is proved.

The last inequality (2.1) is obvious by the elementary fact

$$\alpha\beta \le \frac{1}{4}(\alpha+\beta)^2, \quad \alpha, \beta \in R$$

in choosing $\alpha = R - 1$, $\beta = 1 - r$. ∎

COROLLARY 2.2 (Dragomir et al., 2001 [**65**]). *Let f be as in Theorem 2.1. If $\varepsilon > 0$ and*

$$0 \le R - r \le 4 \cdot \sqrt{\frac{\varepsilon}{\|f''\|_\infty}} \tag{2.9}$$

then

$$\left| I_f(p,q) - f(1) - \sum_{i=1}^{n} (p_i - q_i) f\left(\frac{p_i + q_i}{2q_i}\right) \right| \le \varepsilon.$$

The above theorem has the following corollary which emphasizes better the approximation aspect of the problem for distribution p and q which are close in a certain sense.

COROLLARY 2.3 (Dragomir et al., 2001 [**65**]). *Let $f : [0,2] \to R$ be so that the derivative $f' : [0,2] \to R$ is absolutely continuous and $f'' \in L_\infty[0,2]$. If $\eta \in (0,1)$ and $p(\eta)$, $q(\eta)$ are so that*

$$\left| \frac{p_i(\eta)}{q_i(\eta)} - 1 \right| \le \eta \text{ for all } i \in \{1, ..., n\}, \tag{2.10}$$

then

$$I_f(p(\eta), q(\eta)) = f(1) + \sum_{i=1}^{n} (p_i(\eta) - q_i(\eta)) f'\left(\frac{p_i(\eta) + q_i(\eta)}{2q_i(\eta)}\right) + R_f(p,q,\eta) \tag{2.11}$$

and the remainder $R_f(p,q,\eta)$ satisfies the estimate

$$R_f(p,q,\eta) \le \frac{1}{4} \left\| f'' \right\|_\infty D_{\chi^2}(p(\eta), q(\eta)) \le \frac{1}{4} \left\| f'' \right\|_\infty \eta^2.$$

If we have more information about the generating function f, for example, if we know that the second derivative f'' is absolutely continuous on $[r, R]$ and $f''' \in L_\infty[r, R]$, then we can state the following theorem as well.

THEOREM 2.4 (Dragomir et al., 2001 [**65**]). *If $f : [a,b] \to \mathbb{R}$ be a mapping such that $f''' \in L_\infty[r,R]$, then we have the following:*

$$\left| I_f(p,q) - f(1) - \sum_{i=1}^{n}(p_i - q_i) f'\left(\frac{p_i+q_i}{2q_i}\right) \right| \le \frac{1}{24}\left\| f''' \right\|_\infty D_{|\chi|^3}(p,q) \le \frac{1}{24}\left\| f''' \right\|_\infty (R-r)^3 \tag{2.12}$$

where

$$D_{|\chi|^3}(p,q) := \sum_{i=1}^{n} \frac{|p_i - q_i|^3}{q_i^2},$$

is the absolute Chi-Cube divergence of p, q.

PROOF. We know the following mid-point inequality arising in Numerical Integration

$$\left| g\left(\frac{a+b}{2}\right)(b-a) - \int_a^b g(t)\,dt \right| \le \frac{1}{24}(b-a)^3 \left\| g'' \right\|_\infty \tag{2.13}$$

provided $g'' \in L_\infty[a,b]$.

Using the same argument as in Theorem 2.1, we can state that

$$\left| f(x) - f(1) - (x-1) f'\left(\frac{1+x}{2}\right) \right| \le \frac{1}{24}|x-1|^3 \left\| f''' \right\|_\infty \tag{2.14}$$

for all $x \in [r,R]$ (we know $r \le 1 \le R$).

Now, choosing $x = \frac{p_i}{q_i} \in [r,R]$ and doing as in the proof of Theorem 2.1, we deduce the first inequality in (2.12).

To prove the second part of (2.12), we take into account that

$$\left| \frac{p_i}{q_i} - 1 \right| \le R - r \text{ for all } i \in \{1,...,n\}.$$

Consequently,

$$D_{|\chi|^3}(p,q) = \sum_{i=1}^{n} q_i \left| \frac{p_i}{q_i} - 1 \right|^3 \le (R-r)^3 \sum_{i=1}^{n} q_i = (R-r)^3$$

and the theorem is then proved. ∎

COROLLARY 2.5 (Dragomir et al., 2001 [**65**]). *Let f be as in Theorem 2.4. It $\varepsilon > 0$ and*

$$0 \le R - r \le 2 \cdot \sqrt[3]{\frac{3\varepsilon}{\|f'''\|_\infty}} \tag{2.15}$$

then

$$\left| I_f(p,q) - f(1) - \sum_{i=1}^{n}(p_i - q_i) f'\left(\frac{p_i+q_i}{2q_i}\right) \right| \le \varepsilon.$$

Also, the following approximation result holds.

COROLLARY 2.6 (Dragomir et al., 2001 [**65**]). *Let $f : [0,2] \to R$ be so that $f''' \in L_\infty[0,2]$. If $\eta \in (0,1)$ and $p(\eta)$, $q(\eta)$ are so that (2.10) holds, then we have the representation (2.11) and the remainder $R_f(p,q,\eta)$ satisfies the estimate:*

$$|R_f(p,q,\eta)| \le \frac{1}{3}\left\|f'''\right\|_\infty \eta^3.$$

3. SOME PARTICULAR INEQUALITIES

1. Consider the mapping $f : (0,\infty) \to R$, $f(u) = u \ln u$. Then we have

$$I_f(p,q) = D(p,q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} \quad \text{(Kulback-Leibler distance)} \tag{3.1}$$

$$\begin{aligned} \sum_{i=1}^{n}(p_i - q_i) f'\left(\frac{p_i+q_i}{2q_i}\right) &= \sum_{i=1}^{n}(p_i - q_i)\left[\ln\left(\frac{p_i+q_i}{2q_i}\right)+1\right] \\ &= \sum_{i=1}^{n}(p_i - q_i)\ln\left(\frac{p_i+q_i}{2q_i}\right). \end{aligned} \tag{3.2}$$

As $f''(u) = \frac{1}{u}$, we have

$$\left\|f''\right\|_\infty = \sup_{u\in[r,R]}\left|f''(u)\right| = \frac{1}{r}. \tag{3.3}$$

Consequently, by (2.1), we can state the inequality:

$$\begin{aligned} &\left|D(p,q) - \sum_{i=1}^{n}(p_i - q_i)\ln\left(\frac{p_i+q_i}{2q_i}\right)\right| \\ &\le \frac{1}{4r} D_{\chi^2}(p,q) \le \frac{1}{4r}(R-1)(1-r) \le \frac{1}{16r}(R-r)^2. \end{aligned} \tag{3.4}$$

2. Consider the mapping $f : (0,\infty) \to R$, $f(u) = \ln u$. Then we have

$$I_f(p,q) = \sum_{i=1}^{n} p_i \ln \frac{p_i}{q_i} = -D(q,p), \tag{3.5}$$

$$\begin{aligned} \sum_{i=1}^{n}(p_i - q_i) f'\left(\frac{p_i+q_i}{2q_i}\right) &= \sum_{i=1}^{n}(p_i - q_i)\frac{2q_i}{p_i+q_i} \\ &= 2\sum_{i=1}^{n} q_i \cdot \frac{p_i - q_i}{p_i+q_i}. \end{aligned} \tag{3.6}$$

As $f''(u) = -\frac{1}{u^2}$, we have

$$\left\|f''\right\|_\infty = \frac{1}{r^2}. \tag{3.7}$$

Consequently, by (2.1), we can state the inequality

$$\begin{aligned} &\left|D(q,p) - 2\sum_{i=1}^{n} q_i\left(\frac{q_i - p_i}{p_i+q_i}\right)\right| \\ &\le \frac{1}{4r^2} D_{\chi^2}(p,q) \le \frac{1}{4r^2}(R-1)(1-r) \le \frac{1}{16}\left(\frac{R}{r}-1\right)^2. \end{aligned} \tag{3.8}$$

3. Consider the mapping $f : (0, \infty) \to R$, $f(u) = \frac{1}{2}\left(\sqrt{u} - 1\right)^2$. Then we have

$$I_f(p, q) = \frac{1}{2}\sum_{i=1}^{n} q_i \left(\sqrt{\frac{p_i}{q_i}} - 1\right)^2 = h^2(p, q), \text{ (Hellinger discrimination)} \tag{3.9}$$

$$\begin{aligned} \sum_{i=1}^{n} (p_i - q_i) f'\left(\frac{p_i + q_i}{2q_i}\right) &= \sum_{i=1}^{n} (p_i - q_i) \frac{\sqrt{\frac{p_i+q_i}{2q_i}} - 1}{2\sqrt{\frac{p_i+q_i}{2q_i}}} \\ &= \sum_{i=1}^{n} (p_i - q_i) \frac{\sqrt{p_i + q_i} - \sqrt{2q_i}}{2\sqrt{p_i + q_i}} \\ &= \frac{1}{2}\sum_{i=1}^{n} (p_i - q_i) \frac{\sqrt{p_i + q_i}\left(\sqrt{p_i + q_i} - \sqrt{2q_i}\right)}{p_i + q_i}. \end{aligned} \tag{3.10}$$

As $f''(u) = \frac{1}{4u^{2/3}}$, $u \in (0, \infty)$, then

$$\left\|f''\right\|_\infty = \sup_{u\in[r,R]} \left|f''(u)\right| = \frac{1}{4r^{2/3}}. \tag{3.11}$$

Using the inequality (2.1), we may state the inequality

$$\begin{aligned} &\left|h^2(p, q) - \frac{1}{2}\sum_{i=1}^{n} \frac{p_i - q_i}{p_i + q_i}\left[\sqrt{p_i + q_i}\left(\sqrt{p_i + q_i} - \sqrt{2q_i}\right)\right]\right| \\ &\leq \frac{1}{16r^{2/3}} D_{\chi^2}(p, q) \leq \frac{1}{16r^{2/3}}(R - 1)(1 - r) \leq \frac{1}{64r^{2/3}}(R - r)^2. \end{aligned} \tag{3.12}$$

4. Consider the mapping $f : (0, \infty) \to R$, $f(u) = u^\alpha, \alpha > 1, u \in [0, \infty)$. Then

$$I_f(p, q) = \sum_{i=1}^{n} p_i^\alpha q_i^{1-\alpha} = R_\alpha(p, q), \text{ (Renyi } \alpha\text{-order distance)} \tag{3.13}$$

$$\begin{aligned} \sum_{i=1}^{n} (p_i - q_i) f'\left(\frac{p_i + q_i}{2q_i}\right) &= \sum_{i=1}^{n} (p_i - q_i) \alpha\left(\frac{p_i + q_i}{2q_i}\right)^{\alpha-1} \\ &= \frac{\alpha}{2^{\alpha-1}}\sum_{i=1}^{n} (p_i - q_i)\left(\frac{p_i + q_i}{q_i}\right)^{\alpha-1}, \end{aligned} \tag{3.14}$$

and as $f''(u) = \alpha(\alpha - 1)u^{\alpha-2}$, we have

$$\left\|f''\right\|_\infty = \delta_\alpha(r, R) = \begin{cases} \alpha(\alpha - 1)R^{\alpha-2} & \text{if } 2 < \alpha < \infty \\ \alpha(\alpha - 1)r^{\alpha-2} & \text{if } 1 < \alpha < 2. \end{cases} \tag{3.15}$$

Using (2.1), we may state the inequality

$$\left|R_\alpha(p,q) - \frac{\alpha}{2^{\alpha-1}}\sum_{i=1}^{n}(p_i-q_i)\left(\frac{p_i}{q_i}+1\right)^{\alpha-1}\right| \tag{3.16}$$
$$\leq \frac{1}{4}\delta_\alpha(r,R)D_{\chi^2}(p,q)$$
$$\leq \frac{1}{4}\times\begin{cases}\alpha(\alpha-1)R^{\alpha-2}(R-1)(1-r) & \text{if } 2<\alpha<\infty\\ \alpha(\alpha-1)r^{\alpha-2}(R-1)(1-r) & \text{if } 1<\alpha<2.\end{cases}$$

Using Theorem 2.4, we may point out other bounds which may be more useful in application.

5. If we reconsider the mapping $f:(0,\infty)\to R$, $f(u)=u\ln u$. Then $f^{'''}(u)=-\frac{1}{u^2}$, $\left\|f^{'''}\right\|_\infty=\frac{1}{r^2}$ and then, by (2.12) we can state the inequality

$$\left|D(p,q)-\sum_{i=1}^{n}(p_i-q_i)\ln\left(\frac{p_i+q_i}{2q_i}\right)\right| \tag{3.17}$$
$$\leq \frac{1}{24r^2}D_{|\chi|^3}(p,q)\leq\frac{1}{24r^2}(R-r)^3.$$

6. In the case when $f(u)=\ln u$, we have obviously $f^{'''}(u)=\frac{2}{3u^3}$, $u\in[r,R]$ and then $\left\|f^{'''}\right\|_\infty=\frac{2}{3r^3}$, and then by (2.12), we can state the inequality

$$\left|D(q,p)-2\sum_{i=1}^{n}q_i\left(\frac{q_i-p_i}{p_i+q_i}\right)\right|\leq\frac{1}{36r^3}D_{|\chi|^3}(p,q)\leq\frac{1}{36}\left(\frac{R}{r}-1\right)^3. \tag{3.18}$$

7. When $f:(0,\infty)\to R$, $f(u)=\frac{1}{2}\left(\sqrt{u}-1\right)^2$, we have $f^{'''}(u)=-\frac{1}{6}\cdot u^{-5/3}$, $u\in[r,R]$ and then $\left\|f^{'''}\right\|_\infty=\frac{1}{6}\cdot\frac{1}{r^{5/3}}$ Consequently, by (2.12) we can state that

$$\left|h^2(q,p)-\frac{1}{2}\sum_{i=1}^{n}\frac{p_i-q_i}{p_i+q_i}\left[\sqrt{p_i+q_i}\left(\sqrt{p_i+q_i}-\sqrt{2q_i}\right)\right]\right| \tag{3.19}$$
$$\leq\frac{1}{144r^{5/3}}D_{|\chi|^3}(p,q)\leq\frac{1}{144r^{5/3}}(R-r)^3.$$

8. Finally, if $f:(0,\infty)\to R$, $f(u)=u^\alpha$, $\alpha>1$ we have $f^{'''}(u)=\alpha(\alpha-1)(\alpha-2)u^{\alpha-3}$ and then

$$\left\|f^{'''}\right\|_\infty=\eta_\alpha(r,R)=\begin{cases}\alpha(\alpha-1)(\alpha-2)R^{\alpha-3} & \text{if } 3<\alpha<\infty\\ \alpha(\alpha-1)(\alpha-2)r^{\alpha-3} & \text{if } 1<\alpha<3.\end{cases}$$

Consequently, using (2.12), we may state

$$\left|R_\alpha(p,q)-\frac{\alpha}{2^{\alpha-1}}\sum_{i=1}^{n}(p_i-q_i)\left(\frac{p_i}{q_i}+1\right)^{\alpha-1}\right| \tag{3.20}$$
$$\leq\frac{1}{24}\eta_\alpha(r,R)D_{|\chi|^3}(p,q)$$
$$\leq\frac{1}{24}\times\begin{cases}\alpha(\alpha-1)(\alpha-2)R^{\alpha-2}(R-r)^3 & \text{if } 3<\alpha<\infty\\ \alpha(\alpha-1)(\alpha-2)r^{\alpha-2}(R-r)^3 & \text{if } 1<\alpha<3.\end{cases}$$

4. APPLICATION FOR MUTUAL INFORMATION

We consider mutual information, which is a measure of the amount of information that one random variable contains about another random variable. It is the reduction of uncertainty of one variable due to the knowledge of the other.

DEFINITION 4.1. Consider two random variables X and Y with a joint probability mass function $t(x,y)$ and marginal probability mass function $p(x)$, $x \in \mathcal{X}$ and $q(y)$, $y \in \mathcal{Y}$. The mutual information is the relative entropy between the joint distribution and the product distribution, i.e.,

$$I(X;Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} t(x,y) \log \left[\frac{t(x,y)}{p(x)p(y)} \right] = D(t(x,y), p(x)q(y)) \tag{4.1}$$

where $D(.,.)$ is the Kullback-Leibler distance.

We assume in what follows that

$$s \le \frac{t(x,y)}{p(x)q(y)} \le S(x,y) \text{ for all } (x,y) \in \mathcal{X} \times \mathcal{Y}. \tag{4.2}$$

It is obvious that $s \le 1 \le S$.

We also define the mutual information on Chi-Square sum

$$I_{\chi^2}(X;Y) = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \frac{t^2(x,y)}{p(x)q(y)} - 1. \tag{4.3}$$

Using the inequality (3.4), we can state the following proposition.

PROPOSITION 4.1 (Dragomir et al., 2001 [**65**]). *Under the above assumption, we have the inequality*

$$\left| I(X;Y) - \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (t(x,y) - p(x)q(y)) \ln \left[\frac{t(x,y) + p(x)q(y)}{2p(x)p(y)} \right] \right|$$
$$\le \frac{1}{4s} I_{\chi^2}(X;Y) \le \frac{1}{4s}(S-1)(1-s) \le \frac{1}{16s}(S-s)^2.$$

A similar result can be stated if we apply the inequality (3.8). We omit the details.

5. BASIC THEOREMS

Racall the Iyengar inequality [**111**], which states the following.

THEOREM 5.1 (Iyengar inequality). *Suppose* $g : [a,b] \to R$ *is absolutely continuous on* $[a,b]$ *and* $g' : [a,b] \to R$ *is essentially bounded, that is,* $g' \in L_\infty[a,b]$. *Then*

$$\left| \int_a^b g(t)dt - \frac{1}{2}(b-a)[g(a) + g(b)] \right|$$
$$\le \frac{1}{4} \left\| g' \right\|_\infty (b-a)^2 - \frac{1}{4 \left\| g' \right\|_\infty} [g(b) - g(a)]^2.$$

Suppose that there exist distinct real numbers r, R with

$$0 < r \le p_i/q_i \le R < \infty \text{ for all } i \in \{1, ..., n\}. \tag{5.1}$$

We assume (5.1) throughout without further comment and keep the assumption that r, R are distinct, which is easily seen to entail that $r < 1 < R$. Suppose $f^\dagger : [r,R] \to \mathbf{R}$ is given by

$$f^\dagger(u) := f(1) + \frac{u-1}{2} f'(u).$$

We shall also make use of Proposition 1 of [**65**], which provides the following.

PROPOSITION 5.2. *Suppose that (5.1) is satisfied with $r < R$ and that $m \geq 1$. Then*

$$D_{|\chi|^m}(p,q) \leq \frac{(R-1)(1-r)}{R-r}\left[(1-r)^{m-1} + (R-1)^{m-1}\right] \leq \left(\frac{R-r}{2}\right)^m.$$

The first inequality is an equality if and only if p, q form a boundary pair with respect to r and R, that is, for each i either $p_i/q_i = r$ or $p_i/q_i = R$. The second inequality is an equality if and only if $R + r = 2$, that is, r and R are equidistant from unity.

We now proceed to our first basic result. We shall make use of $f_0 : [r, R] \to \mathbf{R}$ given by

$$f_0(u) := \left[f'(u) - f'(1)\right]^2.$$

THEOREM 5.3 (Dragomir et al.,2001 [**67**]). *Suppose $f : [r, R] \to \mathbf{R}$ with f' absolutely continuous on $[r, R]$ and $f'' \in L_\infty[r, R]$. Then*

$$\begin{aligned} \left|I_f(p,q) - I_{f^\dagger}(p,q)\right| &\leq \frac{1}{4}\left\|f''\right\|_\infty D_{\chi^2}(p,q) - \frac{1}{4\|f''\|_\infty} I_{f_0}(p,q) \\ &\leq \frac{1}{4}\left\|f''\right\|_\infty D_{\chi^2}(p,q) \leq \frac{1}{4}\left\|f''\right\|_\infty (R-1)(1-r) \\ &\leq \frac{1}{16}\left\|f''\right\|_\infty (R-r)^2. \end{aligned} \tag{5.2}$$

PROOF. We choose $a = 1$, $b = x$ and $g = f'$ in Theorem 5.1 to obtain

$$\begin{aligned} &\left|\int_1^x f'(t)dt - \frac{1}{2}(x-1)\left[f'(1) + f'(x)\right]\right| \\ &\qquad \leq \frac{1}{4}\left\|f''\right\|_\infty (x-1)^2 - \frac{1}{4\|f''\|_\infty} f_0(x), \end{aligned}$$

or equivalently

$$\begin{aligned} &\left|f(x) - f(1) - \frac{1}{2}f'(1)(x-1) - \frac{1}{2}f'(x)(x-1)\right| \\ &\qquad \leq \frac{1}{4}\left\|f''\right\|_\infty (x-1)^2 - \frac{1}{4\|f''\|_\infty} f_0(x) \end{aligned}$$

for all $x \in [r, R]$.

If we choose $x = p_i/q_i$, multiply by q_i and sum over i from 1 to n, we derive the first inequality in (5.2) *via* the extended triangle inequality. The second inequality is immediate, since f_0 and so also I_{f_0} is nonnegative. The remaining inequalities are given by Proposition 5.2 with $m = 2$. ∎

COROLLARY 5.4. *Suppose the assumptions of Theorem 5.3 hold. If $\varepsilon > 0$ and*

$$0 \leq R - r \leq 4 \cdot \sqrt{\varepsilon / \|f''\|_\infty},$$

then

$$\left|I_f(p,q) - I_{f^\dagger}(p,q)\right| \leq \varepsilon.$$

The following corollary emphasizes the approximation aspect of Theorem 5.3 for distributions p and q which are close.

COROLLARY 5.5. *Let* $f : [0,2] \to \mathbf{R}$ *be such that* $f' : [0,2] \to \mathbf{R}$ *is absolutely continuous and* $f'' \in L_\infty[0,2]$. *If* $\eta \in (0,1)$ *and* $p(\eta)$, $q(\eta)$ *satisfy*

$$\left| \frac{p_i(\eta)}{q_i(\eta)} - 1 \right| \le \eta \text{ for all } i \in \{1, ..., n\}, \tag{5.3}$$

then

$$I_f(p(\eta), q(\eta)) = I_{f^\dagger}(p(\eta), q(\eta)) + R_f(p, q, \eta) \tag{5.4}$$

and the remainder $R_f(p,q,\eta)$ *satisfies*

$$R_f(p,q,\eta) \le \frac{1}{4} \left\| f'' \right\|_\infty \eta^2.$$

PROOF. Choose $r = 1 - \eta$, $R = 1 + \eta$ in Theorem 5.3. ∎

Our other basic theorem makes use of the trapezoid inequality

$$\left| \int_a^b g(t)\, dt - \frac{1}{2}(b-a)\left[g(a) + g(b)\right] \right| \le \frac{1}{12}(b-a)^3 \left\| g'' \right\|_\infty \tag{5.5}$$

from numerical integration, which holds provided $g'' \in L_\infty[a,b]$.

THEOREM 5.6 (Dragomir et al.,2001 [**67**]). *If* $f : [r,R] \to \mathbf{R}$ *with* f'' *absolutely continuous on* $[r,R]$ *and* $f''' \in L_\infty[r,R]$, *then*

$$\begin{aligned} \left| I_f(p,q) - I_{f^\dagger}(p,q) \right| &\le \frac{1}{12} \left\| f''' \right\|_\infty D_{|\chi|^3}(p,q) \\ &\le \frac{1}{24} \left\| f''' \right\|_\infty \frac{(R-1)(1-r)}{R-r} \left[(1-r)^2 + (R-1)^2 \right] \\ &\le \frac{1}{96} \left\| f''' \right\|_\infty (R-r)^3. \end{aligned} \tag{5.6}$$

The constants on the right are best–possible.

PROOF. The first inequality is derived from (5.5) along the same lines as the first inequality in Theorem 5.3. The remaing inequalities follow from Proposition 5.2 with $m = 3$.

For $f(u) = |u-1|^3$ we have $I_f(p,q) = D_{|\chi|^3}(p,q)$. Also $f^\dagger(u) = (3/2)f(u)$, so as I is linear in f we have

$$\left| I_f(p,q) - I_{f^\dagger}(p,q) \right| = \frac{1}{2} D_{|\chi|^3}(p,q).$$

Since $\left\| f''' \right\| = 6$, the first inequality in (5.6) is thus an equality for this choice of f and the corresponding constant is best–possible. That the following constants are best–possible is inherited from Proposition 5.2. ∎

COROLLARY 5.7. *Let* f *be as in Theorem 5.6. If* $\varepsilon > 0$ *and*

$$0 \le R - r \le 2 \cdot \sqrt[3]{12\varepsilon / \left\| f''' \right\|_\infty},$$

then

$$\left| I_f(p,q) - I_{f^\dagger}(p,q) \right| \le \varepsilon.$$

Also, the following approximation result holds.

COROLLARY 5.8. *Let* $f : [0,2] \to \mathbf{R}$ *with* f'' *absolutely continuous on* $[0,2]$ *and* $f''' \in L_\infty[0,2]$. *If* $\eta \in (0,1)$ *and* $p(\eta)$, $q(\eta)$ *are such that (5.3) holds, then we have the representation (5.4) and the remainder* $R_f(p,q,\eta)$ *satisfies*

$$|R_f(p,q,\eta)| \leq \frac{1}{12} \left\| f''' \right\|_\infty \eta^3.$$

REMARK 5.1. The last bound in Theorem 5.6 is tighter than that of Theorem 5.3 if

$$\frac{6\left\| f'' \right\|_\infty}{\left\| f''' \right\|_\infty} > R - r,$$

while the reverse is true if

$$\frac{6\left\| f'' \right\|_\infty}{\left\| f''' \right\|_\infty} < R - r.$$

As we shall see, both possibilities can arise in practice. In the examples we consider, Theorem 5.6 gives the better bound when r/R is large and Theorem 5.3 when it is small.

6. ONE– AND TWO–POINT DISTRIBUTIONS

Suppose we wish to obtain the error bound involved in estimating $I_f(p,q)$ by $I_g(p,q)$ for some function g. Since

$$I_f(p,q) - I_g(p,q) = I_{f-g}(p,q),$$

we wish to find $\sup |I_h(p,q)|$, where $h = f - g$ and the supremum is taken over all n–point probability distribution pairs (p,q) satisfying (5.1). In this section we approach this question directly and establish some basic results. We shall find it convenient for clarity and succinctness to adopt the notation $I_{h,n}(p,q)$ and to write $u_i := p_i/q_i$ for $i \in \{1,2,\dots,n\}$ in this and the following section.

PROPOSITION 6.1 (Dragomir et al.,2001 [**67**]). *Let* $n \geq 1$ *and assume* p, q *are* n*–point distributions all of whose components are nonzero.*
(a) There exist k*–point distributions* p^U, q^U*, where* k *takes one of the values 1 or 2 and* p^U, q^U *depend on* p *and* q*, such that*

$$p_i^U/q_i^U \in \{u_1,\dots,u_n\} \text{ for } i = 1,\dots,k$$

and

$$I_{h,k}(p^U,q^U) \geq I_{h,n}(p,q).$$

(b) There exist k*–point distributions* p^L, q^L*, where* k *takes one of the values 1 or 2 and* p^L, q^L *depend on* p *and* q*, such that*

$$p_i^L/q_i^L \in \{u_1,\dots,u_n\} \text{ for } i = 1,\dots,k$$

and

$$I_{h,k}(p^L,q^L) \leq I_{h,n}(p,q).$$

PROOF. Consider the first half of the enunciation. If there are j distinct values $u_{(1)}, u_{(2)}, \dots, u_{(j)}$ $(1 \leq j \leq n)$, then for each such we may sum the associated values q_i to obtain $q_1^{(j)}, q_2^{(j)}, \dots, q_j^{(j)}$. Likewise we derive $p_\ell^{(j)}$ associated with $u_{(\ell)}$ $(1 \leq \ell \leq j)$. We have at once that $p_\ell^{(j)}/q_\ell^{(j)} = u_{(\ell)}$, and that $p^{(j)} = (p_1^{(j)}, \dots, p_j^{(j)})$ and $q^{(j)} = (q_1^{(j)}, \dots, q_j^{(j)})$ are j–point probability distributions for which $I_{h,j}(p^{(j)}, q^{(j)})$ has the same value as $I_{h,n}(p,q)$.

To derive the desired result (a) by mathematical induction, we need to show that if m is such that $3 \leq m \leq j$ then there exist $(m-t)$–point probability distributions $p^{(m-t)}$, $q^{(m-t)}$ (with t equal to 1 or 2) depending on $p^{(m)}$ and $q^{(m)}$ with

$$p_\ell^{(m-t)}/q_\ell^{(m-t)} = u_{(\ell)} \quad (1 \leq \ell \leq m-t)$$

and

$$I_{h,m-t}(p^{(m-t)}, q^{(m-t)}) \geq I_{h,j}(p^{(m)}, q^{(m)}).$$

We may without loss of generality assume that there are at least three distinct values u_i, for otherwise there is nothing to prove.

To achieve the induction, we shall show that such a reduction from m–point support to $(m-t)$–point support can be brought about by replacing the last three components of $p^{(m)}$ by two suitably chosen components and a zero or one suitably chosen component and two zeros, with a corresponding replacement in $q^{(m)}$, the zeros being in the same position or positions.

For notational convenience, put $v_1 := u_{(m-2)}$, $v_2 := u_{(m-1)}$, $v_3 := u_{(m)}$. With relabelling if necessary, we may assume $v_1 < v_2 < v_3$. Likewise we put $p_{m-2}^{(m)} = \rho_1$, $p_{m-1}^{(m)} = \rho_2$, $p_m^{(m)} = \rho_3$ and $q_{m-2}^{(m)} = \sigma_1$, $q_{m-1}^{(m)} = \sigma_2$, $q_m^{(m)} = \sigma_3$, so that $\rho_i/\sigma_i = v_i$ for $i = 1, 2, 3$. Define

$$\lambda := \frac{v_3 - v_2}{v_3 - v_1},$$

so that $0 < \lambda, 1-\lambda < 1$ and $v_2 = \lambda v_1 + (1-\lambda)v_3$. We address in turn three possible cases.

(i) If $h(v_2) \leq \lambda h(v_1) + (1-\lambda)h(v_3)$, then

$$\sum_{i=1}^{3} \sigma_i h(v_i) \leq \sum_{i=1,3} \sigma_i' h(v_i),$$

where $\sigma_1' = \sigma_1 + \lambda\sigma_2$ and $\sigma_3' = \sigma_3 + (1-\lambda)\sigma_2$. Note that $\sum_{i=1}^3 \sigma_i = \sum_{i=1,3} \sigma_i'$. If we define $\rho_i' = \sigma_i' v_i$ for $i = 1, 3$, then $\rho_i'/\sigma_i' = v_i$ ($i = 1, 3$) and

$$\sum_{i=1,3} \rho_i' = \sum_{i=1,3} \sigma_i v_i + \sigma_2[\lambda_1 v_1 + (1-\lambda)v_3] = \sum_{i=1}^{3} \sigma_i v_i = \sum_{i=1}^{3} \rho_i.$$

This shows that the reduction can be effected in this case with $t = 1$.

(ii) Next we suppose

$$h(v_2) > \lambda h(v_1) + (1-\lambda)h(v_3) \tag{6.1}$$

with

$$\sigma_1/\lambda \leq \sigma_3/(1-\lambda). \tag{6.2}$$

Put

$$\sigma_2' := \sigma_2 + \sigma_1/\lambda, \quad \sigma_3' := \sigma_3 - \frac{1-\lambda}{\lambda}\sigma_1.$$

Then $\sigma_2' > 0$ and by (6.2) $\sigma_3' \geq 0$. Further $\sum_{i=2,3} \sigma_i' = \sum_{i=1}^3 \sigma_i$. Also by (6.1),

$$\begin{aligned}\sum_{i=2,3} \sigma_i' h(v_i) &> \frac{\sigma_1}{\lambda}[\lambda h(v_1) + (1-\lambda)h(v_3)] + \sigma_2 h(v_2) + \left[\sigma_3 - \frac{1-\lambda}{\lambda}\sigma_1\right] h(v_3) \\ &= \sum_{i=1}^{3} \sigma_i h(v_i).\end{aligned}$$

If we define $\rho_i' = \sigma_i' v_i$ for $i = 2, 3$, then $\rho_i'/\sigma_i' = v_i$ ($i = 2, 3$) and much as above

$$\sum_{i=2,3} \rho_i' = \sum_{i=1}^{3} \rho_i.$$

If (6.2) holds with equality, then $\sigma_3' = 0$ and the reduction holds with $t = 2$. Otherwise it holds with $t = 1$.

(iii) Finally we have the possibility that (6.1) holds with

$$\sigma_1/\lambda > \sigma_3/(1-\lambda). \tag{6.3}$$

We may argue as in (ii), this time starting with

$$\sigma_1' := \sigma_1 - \frac{\lambda}{1-\lambda}\sigma_3, \quad \sigma_2' := \sigma_2 + \frac{1}{1-\lambda}\sigma_3.$$

Then σ_1', σ_2' are positive and

$$\sum_{i=1,2} \sigma_i' = \sum_{i=1^3} \sigma_i.$$

By (3.1),

$$\begin{aligned}\sum_{i=1,2} \sigma_i' h(v_i) &= \sum_{i=1,2} \sigma_i h(v_i) + \frac{\sigma_3}{1-\lambda}\left[h(v_2) - \lambda h(v_1)\right] \\ &> \sum_{i=1}^{3} \sigma_i h) v_i).\end{aligned}$$

If $\rho_i' = \sigma_i' v_i$ for $i = 1, 2$, then $\rho_i'/\sigma_i' = v_i$ ($i = 1, 2$) and

$$\sum_{i=1,2} \rho_i' = \sum_{i=1,2} \sigma_i' v_i = \sum_{i=1,2} \sigma_i v_i + \frac{\sigma_3}{1-\lambda}\left[v_2 - \lambda v_1\right] = \sum_{i=1}^{3} \sigma_i v_i = \sum_{i=1}^{3} \rho_i.$$

Thus we have a reduction with $t = 1$.

This completes the proof of part (a). Part (b) follows by applying part (a) to the function $-h$. ∎

Suppose h is bounded on $[r, R]$. By letting p_i, q_i tend to zero for different choices of i successively while keeping $p_\ell/q_\ell \in [r, R]$ for all $\ell \in \{1, 2, \ldots, n\}$, we can obtain one– and two–point distributions satisfying (5.1) as limiting cases of n–point distributions. With this convention, the following result is natural.

PROPOSITION 6.2 (Dragomir et al.,2001 [**67**]). *Suppose h is continuous and bounded on $[r, R]$. Then $I_{h,n}$ achieves its supremum and infimum over n–point distributions p, q satisfying (5.1). These are realised by one– or two–point distributions.*

PROOF. The first part is immediate. The second follows *via* Proposition 6.1, by relating extremum–achieving distributions to I_h evaluated at one– or two–point distributions at which I_h is dominating (in the supremum case) or dominated (in the infimum case). ∎

COROLLARY 6.3. *The supremum and infimum of I_h subject to (5.1) take one of the forms*

$$qh(u) + (1-q)h\left(\frac{1-qu}{1-q}\right), \quad h(1), \tag{6.4}$$

where $u \in [r, R]$.

PROOF. The first form in (6.4) is immediate, since if $p/q = u$, then

$$\frac{1-p}{1-q} = \frac{1-qu}{1-q}.$$

The second is trivial, since $p/q = 1$ when p and q are one–point distributions. ∎

The ideas of this section provide the means for obtaining tight bounds on $|I_h|$ in terms of r, R. This we pursue in the following section. In practice the calculations can be quite intricate even when h has a relatively simple functional form, although they are suited to efficient numerical implementation, such as by bifurcation search.

Finite–point distributions have a special role in extremal theory. For a general discussion, the reader is referred to [**93**].

7. EVALUATING EXTREMA

We now draw together the ideas of the preceding section to codify the treatment of some broad classes of function h. For notational convenience we introduce

$$F(x,y) := \frac{y-1}{y-x}h(x) + \frac{1-x}{y-x}h(y),$$

which gives the value of I_h for two–point distributions p, q with support at $u = x, y$. We assume throughout that (5.1) applies and that $x, y \in [r, R]$.

THEOREM 7.1 (Dragomir et al.,2001 [**67**]). *Suppose r_T, R_T satisfy $r \le r_T < 1 < R_T \le R$ and the line joining $(r_T, h(r_T))$ to $(R_T, h(R_T))$ lies strictly above the graph of $h(u)$ for $u \in [r, R] \setminus \{r_T, R_T\}$. Then*

$$\sup\ I_h(p,q) = F(r_T, R_T).$$

PROOF. Put $v_1 = r_T$, $v_3 = R_T$ and suppose if possible $u = v_2 \in (r_T, R_T)$ is in the support of one– or two–point distributions p_0, q_0 for which I_f realises its supremum. By assumption

$$h(v_2) < \lambda h(v_1) + (1-\lambda)h(v_3)$$

in the notation of Proposition 6.1 (a) case (i). The argument of case (i) shows that $I_f(p_0, q_0)$ experiences a strict increase if p_0, q_0 are modified by a suitable redistribution of probability mass from v_2 to v_1 and v_3, a contradiction to the extremality of $I_f(p_0, q_0)$. Thus the support of p_0 and q_0 must have empty intersection with (r_T, R_T). There is nothing more to prove if $r_T = r$ and $R_T = R$. In any case, we see that since $1 \in (r_T, R_T)$, p_0 and q_0 must have two–point support.

If $R_T < R$, suppose if possible p_0 and q_0 have a point of support $r_0 \in (R_T, R]$. Then $(R_T, h(R_T))$ lies above the chord joining $(r_T, h(r_T))$ to $(r_0, h(r_0))$, so we may derive a contradiction by the construction of case (ii) or case (iii) of Proposition 6.1 (a). Since p_0 and q_0 must have a point of support greater than unity, that point must therefore be $u = R_T$. A similar argument show that the support point less than unity must be at $u = r_T$. ∎

By taking $-h$ in place of h in the preceding theorem, we derive the following corresponding theorem for infima.

THEOREM 7.2 (Dragomir et al.,2001 [**67**]). *Suppose r_S, R_S satisfy $r \le r_S < 1 < R_S \le R$ and the line joining $(r_S, h(r_S))$ to $(R_S, h(R_S))$ lies strictly below the graph of $h(u)$ for $u \in [r, R] \setminus \{r_S, R_S\}$. Then*

$$\inf\ I_h(p,q) = F(r_S, R_S).$$

Whenever the above theorems are applicable, we may derive $\sup\ |I_f(p,q)|$ from

$$\sup\ |I_f(p,q)| = \max\left[\sup\ I_f(p,q), -\inf\ I_f(p,q)\right].$$

Some modification to Theorem 7.1 is necessary if $r_T < 1 < R_T$ is violated. This has not been found to occur in the examples we have looked at but could be dealt with on an *ad hoc* basis. For this reason we have not seen fit to strive for further generality in Theorem 7.1 at the cost of complicating it. The same comment applies to Theorem 7.2.

8. EXAMPLES

For the mapping $f : (0,\infty) \to R$ given by $f(u) = u \ln u$, $I_f(p,q)$ becomes the Kulback–Leibler distance

$$D(p,q) := \sum_{i=1}^{n} p_i \ln(u_i).$$

We have

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

$$\sum_{i=1}^{n} q_i f_0(u_i) = \sum_{i=1}^{n} q_i \left[\ln(u_i)\right]^2$$

and

$$\left\|f^{''}\right\|_{\infty} = \sup_{u\in[r,R]}\left|f^{''}(u)\right| = 1/r.$$

Consequently Theorem 5.3 provides

$$\begin{aligned} 0 &\le |D(p,q) - D(q,p)| \\ &\le \frac{1}{2r}D_{\chi^2}(p,q) - \frac{r}{2}\sum_{i=1}^{n} q_i\left[\ln\left(\frac{p_i}{q_i}\right)\right]^2 \\ &\le \frac{1}{2r}D_{\chi^2}(p,q) \le \frac{1}{2r}(R-1)(1-r) \le \frac{1}{8r}(R-r)^2 \end{aligned} \tag{8.1}$$

which measures the asymmetry of the Kullback–Leibler distance.

A simple calculation gives

$$\left\|f^{'''}\right\|_{\infty} = 1/r^2,$$

so that Theorem 5.6 gives the bounds

$$0 \le |D(p,q) - D(q,p)| \le \frac{1}{6r^2}D_{|\chi|^3}(p,q) \le \frac{1}{48r^2}(R-r)^3.$$

The last bound here can be seen to be strictly better than that in (8.1) if $r > R/7$.

Tight bounds for $\frac{1}{2}|D(p,q) - D(q,p)|$ involving only r, R can be derived using the ideas of the previous section. We have

$$h(u) = \frac{u+1}{2}\left[\ln u - (u-1)\right],$$

so that

$$h^{'}(u) = \frac{1}{2}\left[\ln u + \frac{1}{u}\right] \text{ and } h^{''}(u) = \frac{u-1}{2u^2}.$$

Thus $h^{'}(u) > 0$ for $u \geq 1$. By the elementary inequality

$$\frac{1}{u} + \ln u > 1 \text{ for } 0 < u < 1,$$

we have that $h^{'}(u) > 0$ holds for $0 < u < 1$ as well and so h is strictly increasing for all $u < 0$. Also h is strictly concave for $0 < u < 1$ and strictly convex for $u > 1$.

This falls within the scope of Theorems 7.1 and 7.2. If r_1 is the demonstrably unique value of u less than unity at which the tangent to the graph of h passes through $(R, h(R))$, then we may choose $r_T = \max(r, r_1)$. If R_1 is the demonstrably unique value of u greater than unity at which the tangent to the graph of h passes through $(r, h(r))$, then we may choose $R_T = \min(R, R_1)$.

Now consider the mapping $f : (0, \infty) \to R$ given by $f(u) = \ln u$. We have

$$I_f(p,q) = \sum_{i=1}^{n} q_i \ln(u_i) = -D(q,p),$$

$$\sum_{i=1}^{n} f^{'}(u_i)(p_i - q_i) = \sum_{i=1}^{n} \frac{q_i}{p_i}(p_i - q_i) = 1 - \sum_{i=1}^{n} \frac{q_i^2}{p_i} = -D_{\chi^2}(q,p),$$

$$\sum_{i=1}^{n} q_i f_0(u_i) = \sum_{i=1}^{n} q_i \left(\frac{q_i}{p_i} - 1\right)^2 = \sum_{i=1}^{n} \frac{q_i(q_i - p_i)^2}{p_i^2}$$

and $\left\|f^{''}\right\|_{\infty} = 1/r^2$.

Consequently, by (5.2), we have

$$\begin{aligned} 0 &\leq \left| D(q,p) + \frac{1}{2} D_{\chi^2}(q,p) \right| \\ &\leq \frac{1}{4r^2} D_{\chi^2}(p,q) - \frac{r^2}{4} \sum_{i=1}^{n} \frac{q_i(q_i - p_i)^2}{p_i^2} \\ &\leq \frac{1}{4r^2} D_{\chi^2}(p,q) \leq \frac{1}{4r^2}(R-1)(1-r) \leq \frac{1}{16r^2}(R-r)^2. \end{aligned} \tag{8.2}$$

A simple calculation shows that $\left\|f^{'''}\right\|_{\infty} = 2/r^3$, so that by Theorem 5.6

$$0 \leq \left[D(q,p) + \frac{1}{2} D_{\chi^2}(q,p) \right] \leq \frac{1}{6r^3} D_{|\chi|^3}(p,q) \leq \frac{1}{48}\left(\frac{R}{r} - 1\right)^3.$$

The last bound here can be seen to be better than that in (8.2) if $r > R/4$.

Again we may obtain a tight bound for $|D(q,p) - \frac{1}{2} D_{\chi^2}(q,p)|$ in terms of r, R alone by use of Theorems 7.1 and 7.2. We have

$$h(u) = \ln u - \frac{u-1}{2u},$$

so that

$$h^{'}(u) = \frac{2u-1}{2u^2} \text{ and } h^{''}(u) = \frac{1-u}{u^3}.$$

Thus h is decreasing for $0 < u < 1/2$ and increasing for $u > 1/2$. Further it is strictly convex for $0 < u < 1$ and strictly concave for $u > 1$, and so is quasiconvex. We may define r_T and R_T exactly as in the previous example.

Finally suppose $f : (0, \infty) \to R$ is given by $f(u) = \frac{1}{2}\left(\sqrt{u} - 1\right)^2$. Then $I_f(p,q)$ becomes the Hellinger discrimination

$$h^2(p,q) := \frac{1}{2} \sum_{i=1}^{n} q_i \left(\sqrt{u_i} - 1\right)^2.$$

We have

$$\sum_{i=1}^{n} f'\left(\frac{p_i}{q_i}\right)(p_i - q_i) = \frac{1}{2}\sum_{i=1}^{n}(q_i - p_i)\frac{\sqrt{q_i}}{\sqrt{p_i}},$$

$$\sum_{i=1}^{n} q_i f_0(u_i) = \sum_{i=1}^{n}\frac{q_i}{p_i}(\sqrt{p_i} - \sqrt{q_i})^2$$

and

$$\left\|f''\right\|_\infty = \sup_{u\in[r,R]}\left|f''(u)\right| = \frac{1}{4r^{3/2}}.$$

Consequently Theorem 5.3 provides

$$\begin{aligned}\left|h^2(p,q) - \frac{1}{4}\sum_{i=1}^{n}(q_i - p_i)\sqrt{\frac{q_i}{p_i}}\right| &\le \frac{1}{16r^{3/2}}D_{\chi^2}(p,q) - r^{3/2}\sum_{i=1}^{n}\frac{q_i}{p_i}\left(\sqrt{p_i} - \sqrt{q_i}\right)^2 \\ &\le \frac{1}{16r^{3/2}}D_{\chi^2}(p,q) \le \frac{1}{16r^{3/2}}(R-1)(1-r) \\ &\le \frac{1}{64r^{3/2}}(R-r)^2.\end{aligned}$$

Also, as $f'''(u) = -\frac{3}{8}u^{-5/2}$, we have $\left\|f'''\right\|_\infty = \sup_{u\in[r,R]}\left|f'''(u)\right| = \frac{3}{8}r^{-5/2}$, and Theorem 5.6 gives

$$\left|h^2(p,q) - \frac{1}{4}\sum_{i=1}^{n}(q_i - p_i)\sqrt{\frac{q_i}{p_i}}\right| \le \frac{1}{32r^{5/2}}D_{|\chi|^3}(p,q) \le \frac{1}{256r^{5/2}}(R-r)^3. \tag{8.3}$$

The largest bound here is better than the largest provided by Theorem 5.3 provided $r > R/5$.

The use of Theorems 7.1 and 7.2 to obtain an absolute upper bound for the left–hand side of the first inequality in (8.3) is more complicated than in the previous examples. We have

$$h(u) = \frac{\left(\sqrt{u} - 1\right)^3}{4\sqrt{u}}.$$

Hence

$$h'(u) = \frac{(\sqrt{u} - 1)^2}{8u^{3/2}}\left[2\sqrt{u} + 1\right] \text{ and } h''(u) = \frac{\sqrt{u} - 1}{16u^3}\left[2 + 3u^{1/2} + 3u - 6u^{3/2}\right].$$

Thus h is strictly increasing for $u > 0$. It is strictly concave for $0 < u < 1$, strictly convex for $1 < u < u_0$ and strictly concave for $u > u_0$, where u_0 is the unique zero exceeding unity of the cubic polynomial $2 + 3x + 3x^2 - 6x^3$.

There exist a unique pair of points $(r_1, h(r_1))$, $(R_1, h(R_1))$ with $r_1 < 1 < R_1$ at which the graph of h has a common tangent which lies above the graph for all $u > 0$ except at the two osculating points. If $r \le r_1 < R_1 \le R$, we may take $r_T = r_1$ and $R_T = R_1$. Suppose $r_1 < r$. Then there exists a unique $u = R_2 > 1$ such that the tangent to the graph at $(R_2, h(R_2))$ passes through $(r, h(r))$ and lies above the graph for $r < u < R_2$. We may choose $r_T = r$, $R_T = \min(R_2, R)$.

If the join of $(r, h(r))$ to $(R, h(R))$ lies below the graph for $r < u < R$, or is tangential to the graph at an intermediate point, we may take $r_S = r$, $R_S = R$. Otherwise, $(r, h(r))$ lies on the tangent to the graph at a unique point $(r_2, h(r_2))$ with $1 < r_2 < u_0$ and this tangent meets the graph again at $(r_3, h(r_3))$ with $r_3 > R$. Similarly $(R, h(R))$ lies on the tangent to the graph at a unique point $(R_2, h(R_2))$ with $1 < R_2 < u_0$ and this tangent meets the graph again at $(R_3, h(R_3))$ with $R_3 < r$. At least one of r_2, R_2 is not unity. If $r_2 \ne 1$, we may take $r_S = r$, $R_T = r_2$. If $R_2 \ne 1$, we may take $r_S = R_2$, $R_S = R$.

CHAPTER 14

Deviation of a Function From the Chord and Applications for f-Divergence

1. Introduction

Consider a function $f:[a,b]\to\mathbb{R}$ and assume that it is bounded on $[a,b]$. The chord that connects its end points $A=(a,f(a))$ and $B=(b,f(b))$ has the equation

$$d_f:[a,b]\to\mathbb{R},\ d_f(t)=\frac{1}{b-a}\left[f(a)(b-t)+f(b)(t-a)\right].$$

We introduce the error in approximating the value of the function $f(t)$ by $d_f(t)$ with $t\in[a,b]$ by $\Phi_f(t)$, $i.e.$, $\Phi_f(t)$ is defined by:

$$\Phi_f(t):=\frac{b-t}{b-a}\cdot f(a)+\frac{t-a}{b-a}\cdot f(b)-f(t). \tag{1.1}$$

The main aim of this paper is to provide sharp upper bounds for the absolute value of the difference $\Phi_f(t)$ in each point $t\in[a,b]$ and under various assumptions on the function f or its derivative f'.

Applications for the f*-divergence functional*

$$I_f(\mathbf{p},\mathbf{q}):=\sum_{i=1}^{n}q_i f\left(\frac{p_i}{q_i}\right), \tag{1.2}$$

where $\mathbf{p}=(p_1,\dots,p_n)$, $\mathbf{q}=(q_1,\dots,q_n)$ are positive sequences, that was introduces by Csiszár, as a generalised measure of information, a "distance function" on the set of probability distributions $\mathbb{P}^n$ are also provided.

2. Preliminary Results

The following simple result, which provides a sharp upper bound for the case of bounded functions, has been stated in [**51**] as an intermediate result needed to obtain a Grüss type inequality.

THEOREM 2.1 (Dragomir, 2008 [**53**]). *If* $f:[a,b]\to\mathbb{R}$ *is a bounded function with* $-\infty<m\le f(t)\le M<\infty$ *for any* $t\in[a,b]$, *then*

$$|\Phi_f(t)|\le M-m. \tag{2.1}$$

The multiplicative constant 1 *in front of* $M-m$ *cannot be replaced by a smaller quantity.*

PROOF. For the sake of completeness, we present a short proof.

Since f is bounded, we have $m(b-t)\le(b-t)f(a)\le(b-t)M$, $m(t-a)\le(t-a)f(b)\le(t-a)M$ and $-(b-a)M\le-(b-a)f(t)\le-(b-a)m$, which gives, by addition and division with $b-a$ that

$$-(M-m)\le\frac{(b-t)f(a)+(t-a)f(b)}{b-a}-f(t)\le M-m,$$

for each $t\in[a,b]$, i.e., the desired inequality (2.1) holds.

Now, assume that there exists a constant $C > 0$ such that $|\Phi_f(t)| \leq C(M-m)$ for any f as in the statement of the theorem. Then, for $t = \frac{a+b}{2}$, we should have

$$\left| \frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right) \right| \leq C(M-m). \tag{2.2}$$

If $f : [a,b] \to \mathbb{R}$, $f(t) = \left|t - \frac{a+b}{2}\right|$, then $f(a) = f(b) = \frac{b-a}{2}$, $f\left(\frac{a+b}{2}\right) = 0$, $M = \frac{b-a}{2}$ and $m = 0$ and the inequality (2.2) becomes $\frac{b-a}{2} \leq C \cdot \frac{b-a}{2}$, which implies that $C \geq 1$. ∎

The case of convex functions has been considered in [**52**] in order to prove another Grüss type inequality. The sharpness of the constant has not been analyzed in the earlier paper.

THEOREM 2.2 (Dragomir, 2007 [**52**]). *If $f : [a,b] \to \mathbb{R}$ is a convex function on $[a,b]$, then*

$$0 \leq \Phi_f(t) \leq \frac{(b-t)(t-a)}{b-a}\left[f'_-(b) - f'_+(a)\right] \leq \frac{1}{4}(b-a)\left[f'_-(b) - f'_+(a)\right] \tag{2.3}$$

for any $t \in [a,b]$.

If the lateral derivatives $f'_-(b)$ and $f'_+(a)$ are finite, then the second inequality and the constant $\frac{1}{4}$ are sharp.

PROOF. For the sake of completeness, we present a complete proof of (2.3) below.

Since f is convex, then

$$\frac{t-a}{b-a} \cdot f(b) + \frac{b-t}{b-a} \cdot f(a) \geq f\left[\frac{(b-t)a + (t-a)b}{b-a}\right] = f(t)$$

for any $t \in [a,b]$, i.e., $\Phi(t) \geq 0$ for any $t \in [a,b]$.

If either $f'_-(b)$ or $f'_+(a)$ are infinite, then the last part of (2.3) is obvious.

Suppose that $f'_-(b)$ and $f'_+(a)$ are finite. Then, by the convexity of f we have $f(t) - f(b) \geq f'_-(b)(t-b)$ for any $t \in (a,b)$. If we multiply this inequality with $t - a \geq 0$, we deduce

$$(t-a)f(t) - (t-a)f(b) \geq f'_-(b)(t-b)(t-a), \quad t \in (a,b). \tag{2.4}$$

Similarly, we get

$$(b-t)f(t) - (b-t)f(a) \geq f'_+(a)(t-a)(b-t), \quad t \in (a,b). \tag{2.5}$$

Adding (2.4) to (2.5) and dividing by $b-a$, we deduce

$$f(t) - \frac{(t-a)f(b) + (b-t)f(a)}{b-a} \geq \frac{(b-t)(t-a)}{b-a}\left[f'_-(b) - f'_+(a)\right],$$

for any $t \in (a,b)$, which proves the second inequality for $t \in (a,b)$.

If $t = a$ or $t = b$, the inequality also holds.

Now, assume that (2.3) holds with D and E greater than zero, i.e.,

$$\Phi_f(t) \leq D \cdot \frac{(b-t)(t-a)}{b-a}\left[f'_-(b) - f'_+(a)\right] \leq E(b-a)\left[f'_-(b) - f'_+(a)\right]$$

for any $t \in [a,b]$. If we choose $t = \frac{a+b}{2}$, then we get

$$\begin{aligned} \frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right) &\leq \frac{1}{4}D(b-a)\left[f'_-(b) - f'_+(a)\right] \\ &\leq E(b-a)\left[f'_-(b) - f'_+(a)\right]. \end{aligned} \tag{2.6}$$

Consider $f : [a,b] \to \mathbb{R}$, $f(t) = \left|t - \frac{a+b}{2}\right|$. Then f is convex, $f(a) = f(b) = \frac{b-a}{2}$, $f\left(\frac{a+b}{2}\right) = 0$, $f'_-(b) = 1$, $f'_+(a) = -1$ and by (2.6) we deduce

$$\frac{b-a}{2} \leq \frac{1}{2}D(b-a) \leq 2E(b-a),$$

which implies that $D \geq 1$ and $E \geq \frac{1}{4}$. ∎

3. THE CASE WHEN f IS OF BOUNDED VARIATION

We start with the following representation result:

LEMMA 3.1 (Dragomir, 2008 [**53**]). *If $f : [a,b] \to \mathbb{R}$ is bounded on $[a,b]$ and $Q : [a,b]^2 \to \mathbb{R}$ is defined by*

$$Q(t,s) := \begin{cases} t-b & \text{if } a \leq s \leq t \\ \\ t-a & \text{if } t < s \leq b, \end{cases} \tag{3.1}$$

then we have the representation

$$\Phi_f(t) = \frac{1}{b-a}\int_a^b Q(t,s)\,df(s), \quad t \in [a,b], \tag{3.2}$$

where the integral in (3.2) is taken in the sense of Riemann-Stieltjes.

PROOF. We have:

$$\begin{aligned}\int_a^b Q(t,s)\,df(s) &= \int_a^t (t-b)\,df(s) + \int_t^b (t-a)\,df(s) \\ &= (t-b)\int_a^t df(s) + (t-a)\int_t^b df(s) \\ &= (t-b)[f(t)-f(a)] + (t-a)[f(b)-f(t)] \\ &= (b-a)\Phi_f(t)\end{aligned}$$

and the identity is proved. ∎

The following estimation result holds.

THEOREM 3.2 (Dragomir, 2008 [**53**]). *If $f : [a,b] \to \mathbb{R}$ is of bounded variation, then*

$$\begin{aligned}|\Phi_f(t)| &\leq \left(\frac{b-t}{b-a}\right)\cdot\bigvee_a^t(f) + \left(\frac{t-a}{b-a}\right)\cdot\bigvee_t^b(f) \\ &\leq \begin{cases} \left[\frac{1}{2} + \left|\frac{t-\frac{a+b}{2}}{b-a}\right|\right]\bigvee_a^b(f); \\ \\ \left[\left(\frac{b-t}{b-a}\right)^p + \left(\frac{t-a}{b-a}\right)^p\right]^{\frac{1}{p}}\left[\left(\bigvee_a^t(f)\right)^q + \left(\bigvee_t^b(f)\right)^q\right]^{\frac{1}{q}} \\ \qquad\qquad \text{if } p>1,\ \frac{1}{p}+\frac{1}{q}=1; \\ \\ \frac{1}{2}\bigvee_a^b(f) + \frac{1}{2}\left|\bigvee_a^t(f) - \bigvee_t^b(f)\right|. \end{cases}\end{aligned} \tag{3.3}$$

The first inequality in (3.3) is sharp. The constant $\frac{1}{2}$ is best possible in the first and third branches.

PROOF. We use the fact that for $p : [\alpha,\beta] \to \mathbb{R}$ continuous and $v : [\alpha,\beta] \to \mathbb{R}$ of bounded variation the Riemann-Stieltjes integral $\int_\alpha^\beta p(t)\,dv(t)$ exists and

$$\left|\int_\alpha^\beta p(t)\,dv(t)\right| \leq \sup_{t\in[\alpha,\beta]}|p(t)|\bigvee_\alpha^\beta(v).$$

Then, by the identity (3.2), we have

$$\begin{aligned}|\Phi_f(t)| &\leq \frac{1}{b-a}\left|(t-b)\int_a^t df(s) + (t-a)\int_t^b df(s)\right| \\ &\leq \frac{1}{b-a}\left[(b-t)\left|\int_a^t df(s)\right| + (t-a)\left|\int_t^b df(s)\right|\right] \\ &\leq \frac{1}{b-a}\left[(b-t)\bigvee_a^t(f) + (t-a)\bigvee_t^b(f)\right],\end{aligned}$$

and the first inequality in (3.3) is proved.

Now, by the Hölder inequality, we have

$$(b-t)\bigvee_a^t(f) + (t-a)\bigvee_t^b(f) \leq \begin{cases} \max\{b-t, t-a\}\left[\bigvee_a^t(f) + \bigvee_t^b(f)\right]; \\ \left[(b-t)^p + (t-a)^p\right]^{\frac{1}{p}}\left[\left(\bigvee_a^t(f)\right)^q + \left(\bigvee_t^b(f)\right)^q\right]^{\frac{1}{q}} \\ \qquad \text{if } p>1,\ \frac{1}{p}+\frac{1}{q}=1; \\ (b-t+t-a)\max\left\{\bigvee_a^t(f), \bigvee_t^b(f)\right\}, \end{cases}$$

which produces the last part of (3.3).

For $t = \frac{1}{2}(a+b)$, (3.3) becomes

$$\left|f\left(\frac{a+b}{2}\right) - \frac{f(a)+f(b)}{2}\right| \leq \frac{1}{2}\bigvee_a^b(f).$$

Assume that there exists a constant $A > 0$ such that

$$\left|f\left(\frac{a+b}{2}\right) - \frac{f(a)+f(b)}{2}\right| \leq A\bigvee_a^b(f). \tag{3.4}$$

If in this inequality we choose $f : [a,b] \to \mathbb{R}$, $f(t) = \left|t - \frac{a+b}{2}\right|$, then we deduce $\frac{b-a}{2} \leq A(b-a)$, which implies that $A \geq \frac{1}{2}$. ∎

COROLLARY 3.3. *If $f : [a,b] \to \mathbb{R}$ is L_1-Lipschitzian on $[a,t]$ and L_2-Lipschitzian on $[t,b]$, $L_1, L_2 > 0$, then*

$$|\Phi_f(t)| \leq \frac{(b-t)(t-a)}{b-a}(L_1+L_2) \leq \frac{1}{4}(b-a)(L_1+L_2) \tag{3.5}$$

for any $t \in [a,b]$.

In particular, if f is $L-$Lipschitzian on $[a,b]$, then

$$|\Phi_f(t)| \leq \frac{2(b-t)(t-a)}{b-a}L \leq \frac{1}{2}(b-a)L. \tag{3.6}$$

The constants $\frac{1}{4}, 2$ and $\frac{1}{2}$ are best possible.

The proof is obvious by Theorem 3.2 on taking into account that any $L-$Lipschitzian function is of bounded variation and $\bigvee_a^b(f) \leq (b-a)L$. The sharpness of the constants follows by choosing the function $f : [a,b] \to \mathbb{R}$, $f(t) = \left|t - \frac{a+b}{2}\right|$ which is Lipschitzian with $L = 1$.

COROLLARY 3.4. *If $f:[a,b]\to\mathbb{R}$ is monotonic nondecreasing on $[a,b]$, then*

$$|\Phi_f(t)| \le \left(\frac{b-t}{b-a}\right)[f(t)-f(a)] + \left(\frac{t-a}{b-a}\right)[f(b)-f(t)] \tag{3.7}$$
$$\le \begin{cases} \left[\frac{1}{2}+\left|\frac{t-\frac{a+b}{2}}{b-a}\right|\right][f(b)-f(a)]; \\ \\ \left[\left(\frac{b-t}{b-a}\right)^p+\left(\frac{t-a}{b-a}\right)^p\right]^{\frac{1}{p}}\left[[f(t)-f(a)]^q+[f(b)-f(t)]^q\right]^{\frac{1}{q}} \\ \hfill \text{if } p>1,\ \frac{1}{p}+\frac{1}{q}=1; \\ \\ \frac{1}{2}[f(b)-f(a)]+\frac{1}{2}\left|f(t)-\frac{f(a)+f(b)}{2}\right|. \end{cases}$$

The first inequality and the constant $\frac{1}{2}$ in the first branch of the second inequality are sharp.

The inequality is obvious from (3.3). For $t=\frac{a+b}{2}$, we get in (3.7)

$$\left|f\left(\frac{a+b}{2}\right)-\frac{f(a)+f(b)}{2}\right| \le \frac{1}{2}[f(b)-f(a)]. \tag{3.8}$$

In (3.8), the constant $\frac{1}{2}$ is sharp since for the monotonic nondecreasing function $f:[a,b]\to\mathbb{R}$

$$f(t)=\begin{cases} 0 & \text{if } t\in\left[a,\frac{a+b}{2}\right]; \\ \\ 1 & \text{if } t\in\left(\frac{a+b}{2},b\right], \end{cases}$$

we obtain in both sides of (3.8) the same quantity $\frac{1}{2}$.

4. THE CASE WHEN f IS ABSOLUTELY CONTINUOUS

Now, if $f:[a,b]\to\mathbb{R}$ is absolutely continuous, then f is differentiable almost everywhere and $\int_a^b f'(s)\,ds = f(b)-f(a)$, where the integral is taken in the Lebesgue sense, and we can state the following representation result.

LEMMA 4.1. *If $f:[a,b]\to\mathbb{R}$ is absolutely continuous, then*

$$\Phi_f(t)=\frac{1}{b-a}\int_a^b Q(t,s)\,f'(s)\,ds, \quad t\in[a,b], \tag{4.1}$$

where the integral is in the Lebesgue sense and Q has been defined in (3.1).

The proof is similar to the proof of Lemma 3.1 and the details are omitted.

We define the Lebesgue $p-$norms as follows:

$$\|g\|_{[\alpha,\beta],s} := \begin{cases} ess\sup\limits_{t\in[\alpha,\beta]}|g(t)| & \text{if } s=\infty, \\ \\ \left(\int_\alpha^\beta |g(t)|^s\,dt\right)^{\frac{1}{s}} & \text{if } s\in[1,\infty). \end{cases}$$

The following estimation holds:

THEOREM 4.2 (Dragomir, 2008 [**53**]). *If f is absolutely continuous, then*

$$|\Phi_f(t)| \leq \left(\frac{b-t}{b-a}\right) \cdot \|f'\|_{[a,t],1} + \left(\frac{t-a}{b-a}\right) \cdot \|f'\|_{[t,b],1} \tag{4.2}$$
$$\leq \begin{cases} \frac{(b-t)(t-a)}{b-a} \|f'\|_{[a,t],\infty} & \text{if } f' \in L_\infty[a,b] \\ \\ \frac{(b-t)(t-a)^{\frac{1}{q}}}{b-a} \|f'\|_{[a,t],p} & \text{if } f' \in L_p[a,b],\ p>1,\ \frac{1}{p}+\frac{1}{q}=1 \end{cases}$$
$$+ \begin{cases} \frac{(t-a)(b-t)}{b-a} \|f'\|_{[t,b],\infty} & \text{if } f' \in L_\infty[a,b] \\ \\ \frac{(t-a)(b-t)^{\frac{1}{\beta}}}{b-a} \|f'\|_{[t,b],\alpha} & \text{if } f' \in L_\alpha[a,b],\ \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1 \end{cases}$$

where the second part should be seen as all four possible combinations.

PROOF. The first inequality holds from the representation (4.1) on taking the modulus and applying its properties.

By the integral Hölder inequality, we have

$$\int_a^t |f'(s)|\, ds \leq \begin{cases} (t-a)\, ess \sup\limits_{s\in[a,t]} |f'(s)| & \text{if } f' \in L_\infty[a,b] \\ (t-a)^{\frac{1}{q}} \left(\int_a^t |f'(s)|^p\, ds\right)^{\frac{1}{p}} & \text{if } f' \in L_p[a,b], \\ & p>1,\ \frac{1}{p}+\frac{1}{q}=1 \end{cases}$$

and

$$\int_t^b |f'(s)|\, ds \leq \begin{cases} (b-t)\, ess \sup\limits_{s\in[t,b]} |f'(s)| & \text{if } f' \in L_\infty[a,b] \\ (b-t)^{\frac{1}{q}} \left(\int_t^b |f'(s)|^p\, ds\right)^{\frac{1}{p}} & \text{if } f' \in L_p[a,b], \\ & p>1,\ \frac{1}{p}+\frac{1}{q}=1 \end{cases}$$

which provides the second part of (4.2). ∎

REMARK 4.1. Some particular inequalities of interest are as follows. If $f' \in L_\infty[a,b]$, then

$$|\Phi_f(t)| \leq \frac{(b-t)(t-a)}{b-a}\left[\|f'\|_{[a,t],\infty} + \|f'\|_{[t,b],\infty}\right] \tag{4.3}$$
$$\leq \frac{2(b-t)(t-a)}{b-a}\|f'\|_{[a,b],\infty} \leq \frac{1}{2}(b-a)\|f'\|_{[a,b],\infty},$$

for any $t \in [a,b]$. The first inequality in (4.3) and the constants 2 and $\frac{1}{2}$ are best possible.

If $f' \in L_p[a,b]$, $p>1$, $\frac{1}{p}+\frac{1}{q}=1$, then

$$|\Phi_f(t)| \leq \left[\frac{(b-t)(t-a)}{b-a}\right]^{\frac{1}{q}} \left[\left(\frac{b-t}{b-a}\right)^{\frac{1}{p}} \|f'\|_{[a,t],p} + \left(\frac{t-a}{b-a}\right)^{\frac{1}{p}} \|f'\|_{[t,b],p}\right] \tag{4.4}$$
$$\leq \left[\frac{(b-t)(t-a)}{b-a}\right]^{\frac{1}{q}} \left[\left(\frac{b-t}{b-a}\right)^{\frac{q}{p}} + \left(\frac{t-a}{b-a}\right)^{\frac{q}{p}}\right]^{\frac{1}{q}} \|f'\|_{[a,b],p}$$

for any $t \in [a,b]$.

In particular, for $p=q=2$, we have

$$|\Phi_f(t)| \le \sqrt{\frac{(b-t)(t-a)}{b-a}}\left[\sqrt{\frac{b-t}{b-a}}\cdot\|f'\|_{[a,t],2}+\sqrt{\frac{t-a}{b-a}}\cdot\|f'\|_{[t,b],2}\right] \le \sqrt{\frac{(b-t)(t-a)}{b-a}}\|f'\|_{[a,b],2} \tag{4.5}$$

for any $t\in[a,b]$.

5. The Case when f' is of Bounded Variation

The following representation of the error Φ_f can be stated:

LEMMA 5.1 (Dragomir, 2008 [**53**]). *If $f:[a,b]\to\mathbb{R}$ is absolutely continuous on $[a,b]$ and such that the derivative f' is Riemann integrable on $[a,b]$, then we have the following representation in terms of the Riemann-Stieltjes integral:*

$$\Phi_f(t)=\frac{1}{b-a}\int_a^b K(t,s)\,df'(s),\quad t\in[a,b], \tag{5.1}$$

where the kernel $K:[a,b]^2\to\mathbb{R}$ is given by

$$K(t,s):=\begin{cases}(b-t)(s-a) & \text{if } a\le s\le t\\ (t-a)(b-s) & \text{if } t<s\le b.\end{cases} \tag{5.2}$$

PROOF. Since f' is Riemann integrable on $[a,b]$, it follows that the Riemann-Stieltjes integrals $\int_a^t(s-a)\,df'(s)$ and $\int_t^b(b-s)\,df'(s)$ exist for each $t\in[a,b]$. Now, integrating by parts in the Riemann-Stieltjes integral, we have:

$$\begin{aligned}\int_a^b K(t,s)\,df'(s) &= (b-t)\int_a^t(s-a)\,df'(s)+(t-a)\int_t^b(b-s)\,df'(s)\\ &=(b-t)\left[(s-a)f'(s)\Big|_a^t-\int_a^t f'(s)\,ds\right]+(t-a)\left[(b-s)f'(s)\Big|_t^b-\int_t^b f'(s)\,ds\right]\\ &=(b-t)\left[(t-a)f'(t)-(f(t)-f(a))\right]+(t-a)\left[-(b-t)f'(t)+f(b)-f(t)\right]\\ &=(t-a)\left[f(b)-f(t)\right]-(b-t)\left[f(t)-f(a)\right]=(b-a)\Phi_f(t)\end{aligned}$$

for any $t\in[a,b]$, which provides the desired representation (5.1). ∎

REMARK 5.1. If we define $\Delta_f:(a,b)\to\mathbb{R}$,

$$\Delta_f(t)=\frac{f(b)-f(t)}{b-t}-\frac{f(t)-f(a)}{t-a},$$

then by the above identity (5.1), we have the representation

$$\Delta_f(t)=\frac{1}{(t-a)(b-t)}\int_a^b K(t,s)\,df'(s)=\int_a^b R(t,s)\,df'(s),\quad t\in(a,b), \tag{5.3}$$

where the new kernel $R:(a,b)^2\to\mathbb{R}$ is defined by

$$R(t,s):=\begin{cases}\frac{s-a}{t-a} & \text{if } a<s\le t\\ \frac{b-s}{b-t} & \text{if } t<s<b.\end{cases}$$

We notice that, for $f(s) := \int_a^s g(z)\,dz$, the last equality in (5.3) produces the following identity:

$$\frac{1}{b-t}\int_t^b g(z)\,dz - \frac{1}{t-a}\int_a^t g(z)\,dz = \int_a^b R(t,s)\,dg(s), \tag{5.4}$$

which has been obtained by P. Cerone in [**13**] (see eq. (2.12)).

Notice that, in (5.4), the function g can be Riemann integrable and not only absolutely continuous as assumed in [**13**].

THEOREM 5.2 (Dragomir, 2008 [**53**]). *Assume that* $f : [a,b] \to \mathbb{R}$ *is absolutely continuous on* $[a,b]$ *. If* f' *is of bounded variation on* $[a,b]$ *, then*

$$|\Phi_f(t)| \leq \frac{(t-a)(b-t)}{b-a} \cdot \bigvee_a^b (f') \leq \frac{1}{4}(b-a)\bigvee_a^b (f'), \tag{5.5}$$

where $\bigvee_a^b (f')$ *denotes the total variation of* f' *on* $[a,b]$ *.*

The inequalities are sharp and the constant $\frac{1}{4}$ *is best possible.*

PROOF. It is well known that, if $p : [\alpha,\beta] \to \mathbb{R}$ is continuous and $v : [\alpha,\beta] \to \mathbb{R}$ is of bounded variation, then the Riemann-Stieltjes integral $\int_\alpha^\beta p(s)\,dv(s)$ exists and

$$\left|\int_\alpha^\beta p(s)\,dv(s)\right| \leq \sup_{s\in[\alpha,\beta]} |p(s)| \bigvee_\alpha^\beta (v).$$

Now, utilising the representation (5.1) and the above property, we have

$$\begin{aligned} |\Phi_f(t)| &= \frac{1}{b-a}\left|(b-t)\int_a^t (s-a)\,df'(s) + (t-a)\int_t^b (t-s)\,df'(s)\right| \\ &\leq \frac{1}{b-a}\left[(b-t)\left|\int_a^t (s-a)\,df'(s)\right| + (t-a)\left|\int_t^b (t-s)\,df'(s)\right|\right] \\ &\leq \frac{1}{b-a}\left[(b-t)\bigvee_a^t (f') \sup_{s\in[a,t]}(s-a) + (t-a)\bigvee_t^b (f') \sup_{s\in[t,b]}(t-s)\right] \end{aligned} \tag{5.6}$$

$$= \frac{(t-a)(b-t)}{b-a}\left[\bigvee_a^t (f') + \bigvee_t^b (f')\right] = \frac{(t-a)(b-t)}{b-a}\bigvee_a^b (f').$$

The last part of (5.5) is obvious by the fact that $(t-a)(b-t) \le \frac{1}{4}(b-a)^2$, $t \in [a,b]$.

For the sharpness of the inequalities in (5.5), assume that there exists $F, G > 0$ such that

$$|\Phi_f(t)| \le F \cdot \frac{(t-a)(b-t)}{b-a}\bigvee_a^b (f') \le G(b-a)\bigvee_a^b (f'),$$

with f as in the assumption of the theorem. Then, for $t = \frac{a+b}{2}$, we get

$$\left|\frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right)\right| \le \frac{1}{4}F(b-a)\bigvee_a^b (f') \le G(b-a)\bigvee_a^b (f'). \tag{5.7}$$

Consider the function $f : [a,b] \to \mathbb{R}$, $f(t) = \left|t - \frac{a+b}{2}\right|$. This function is absolutely continuous, $f'(t) = \operatorname{sgn}\left(t - \frac{a+b}{2}\right)$, $t \in [a,b] \setminus \left\{\frac{a+b}{2}\right\}$ and $\bigvee_a^b (f') = 2$. Thus, (5.7) becomes

$$\frac{b-a}{2} \le \frac{1}{2}F(b-a) \le 2G(b-a),$$

which implies that $F \ge 1$ and $G \ge \frac{1}{4}$. ∎

6. THE CASE WHEN f' IS LIPSCHITZIAN

The case when the derivative is a Lipschitzian function provides better accuracy in approximating the function f by the straight line d_f as follows:

THEOREM 6.1 (Dragomir, 2008 [**53**]). *Assume that $f : [a,b] \to \mathbb{R}$ is absolutely continuous on $[a,b]$. If f' is K_1-Lipschitzian on $[a,t]$ and K_2-Lipschitzian on $[t,b]$ $(t \in [a,b])$, then*

$$\begin{aligned} |\Phi_f(t)| &\le \frac{1}{2} \cdot \frac{(t-a)(b-t)}{b-a}\left[(K_1-K_2)t + K_2 b - K_1 a\right] \\ &\le \frac{1}{8} \cdot (b-a)\left[(K_1-K_2)t + K_2 b - K_1 a\right], \quad t \in [a,b]. \end{aligned} \tag{6.1}$$

In particular, if f' is $K-$Lipschitzian on $[a,b]$, then

$$|\Phi_f(t)| \le \frac{1}{2}(b-t)(t-a)K \le \frac{1}{8}(b-a)^2 K, \quad t \in [a,b]. \tag{6.2}$$

The constants $\frac{1}{2}$ and $\frac{1}{8}$ are best possible.

PROOF. We utilize the fact that for an $L-$Lipschitzian function, $p : [\alpha,\beta] \to \mathbb{R}$ and a Riemann integrable function $v : [\alpha,\beta] \to \mathbb{R}$, the Riemann-Stieltjes integral $\int_\alpha^\beta p(s)\,dv(s)$ exists and

$$\left|\int_\alpha^\beta p(s)\,dv(s)\right| \le L\int_\alpha^\beta |p(s)|\,ds.$$

Then we have

$$\left|\int_a^t (s-a)\,df'(s)\right| \le K_1 \cdot \int_a^t (s-a)\,ds = \frac{1}{2}K_1(t-a)^2 \tag{6.3}$$

and

$$\left|\int_t^b (t-s)\,df'(s)\right| \le K_2 \cdot \int_t^b (t-s)\,ds = \frac{1}{2}K_2(b-t)^2. \tag{6.4}$$

Now, on making use of the inequality (5.6) we have, by (6.3) and (6.4), that

$$|\Phi_f(t)| \le \frac{1}{b-a}\left[\frac{1}{2}(b-t)(t-a)^2 \cdot K_1 + \frac{1}{2}(t-a)(b-t)^2 \cdot K_2\right]$$
$$= \frac{1}{2} \cdot \frac{(t-a)(b-t)}{b-a}\left[L_1(t-a) + L_2(b-t)\right],$$

which produces the first inequality in (6.1). The other inequalities are obvious.

To prove the sharpness of the constants in (6.2), let us assume that there exist $H, K > 0$ so that

$$|\Phi_f(t)| \le H(b-t)(t-a)L \le K(b-a)^2 L \tag{6.5}$$

for any $t \in [a,b]$ and f an $L-$Lipschitzian function on $[a,b]$. For $t = \frac{a+b}{2}$ we get from (6.5) that

$$\left|\frac{f(a)+f(b)}{2} - f\left(\frac{a+b}{2}\right)\right| \le \frac{1}{4}HL(b-a)^2 \le LK(b-a)^2. \tag{6.6}$$

Consider $f : [a,b] \to \mathbb{R}$, $f(t) = \frac{1}{2}\left(t - \frac{a+b}{2}\right)^2$. Then $f'(t) = t - \frac{a+b}{2}$ is Lipschitzian with the constant $L = 1$ and (6.6) becomes

$$\frac{1}{8}(b-a)^2 \le \frac{1}{4}H(b-a)^2 \le K(b-a)^2,$$

which implies that $H \ge \frac{1}{2}$ and $K \ge \frac{1}{8}$. ∎

7. The Case when f' is Absolutely Continuous

The following representation result also holds.

Lemma 7.1. *If $f : [a,b] \to \mathbb{R}$ is differentiable and the derivative f' is absolutely continuous, then*

$$\Phi_f(t) = \frac{1}{b-a}\int_a^b K(t,s) f''(s)\, ds \tag{7.1}$$

for any $t \in [a,b]$, where the integral in (7.1) is considered in the Lebesgue sense.

The proof is similar to the one in Lemma 5.1 on integrating by parts in the Lebesgue integral $\int_a^b K(t,s) f''(s)\, ds$. The details are omitted.

Theorem 7.2 (Dragomir, 2008 [**53**]). *If f is as in Lemma 7.1, then*

$$|\Phi_f(t)| \le \frac{(b-t)(t-a)}{b-a} \cdot K(t), \quad t \in [a,b], \tag{7.2}$$

where

$$K(t) := \begin{cases} \|f''\|_{[a,t],1}; & \\ \frac{(t-a)^{1/q}}{(q+1)^{1/q}}\|f''\|_{[a,t],p} & \text{if } p>1,\ \frac{1}{p}+\frac{1}{q}=1, \\ & f'' \in L_p[a,b]; \\ \frac{1}{2}(t-a)\|f''\|_{[a,t],\infty} & \text{if } f'' \in L_\infty[a,b]; \end{cases} \tag{7.3}$$
$$+ \begin{cases} \|f''\|_{[t,b],1} & \\ \frac{(b-t)^{1/\beta}}{(\beta+1)^{1/\beta}}\|f''\|_{[t,b],\alpha} & \text{if } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1, \\ & f'' \in L_\alpha[a,b]; \\ \frac{1}{2}(b-t)\|f''\|_{[t,b],\infty} & \text{if } f'' \in L_\infty[a,b]; \end{cases}$$

and the definition of K should be seen as all 9 possible combinations.

PROOF. We have, by (5.6) that

$$|\Phi_f(t)| \le \frac{1}{b-a}\left[(b-t)\left|\int_a^t (s-a) f''(s)\,ds\right| + (t-a)\left|\int_t^b (t-s) f''(s)\,ds\right|\right] \tag{7.4}$$

for any $t \in [a,b]$.

Utilising Hölder's inequality, we have

$$\begin{aligned}\left|\int_a^t (s-a) f''(s)\,ds\right| &\le \begin{cases} \sup\limits_{s\in[a,t]} (s-a) \int_a^t |f''(s)|\,ds; & \\ \left(\int_a^t (s-a)^q\,ds\right)^{1/q}\left(\int_a^t |f''(s)|^p\,ds\right)^{1/p} & \text{if } p>1,\ \frac{1}{p}+\frac{1}{q}=1, \\ & f''\in L_p[a,b]; \\ ess \sup\limits_{s\in[a,t]} |f''(s)| \int_a^t (s-a)\,ds & \text{if } f''\in L_\infty[a,b]; \end{cases} \\ &= \begin{cases} (t-a)\|f''\|_{[a,t],1}, & \\ \frac{(t-a)^{1+1/q}}{(q+1)^{1/q}}\|f''\|_{[a,t],p} & \text{if } p>1,\ \frac{1}{p}+\frac{1}{q}=1, \\ & f''\in L_p[a,b]; \\ \frac{1}{2}(t-a)^2\|f''\|_{[a,t],\infty} & \text{if } f''\in L_\infty[a,b]; \end{cases}\end{aligned} \tag{7.5}$$

and, similarly,

$$\left|\int_t^b (b-s) f''(s)\,ds\right| \le \begin{cases} (b-t)\|f''\|_{[t,b],1} & \\ \frac{(b-t)^{1+1/\beta}}{(\beta+1)^{1/\beta}}\|f''\|_{[t,b],\alpha} & \text{if } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1, \\ & f''\in L_\alpha[a,b]; \\ \frac{1}{2}(b-t)^2\|f''\|_{[t,b],\infty} & \text{if } f''\in L_\infty[a,b], \end{cases} \tag{7.6}$$

for any $t \in [a,b]$.

Finally, on making use of (7.4) – (7.6), we deduce the desired inequality (7.2). ∎

REMARK 7.1. The inequalities in (7.2) have some instances of interest that are useful in applications. For example, in terms of the sup-norm we have:

$$\begin{aligned}|\Phi_f(t)| &\le \frac{1}{2}\cdot\frac{(b-t)(t-a)}{b-a}\left[(t-a)\|f''\|_{[a,t],\infty} + (b-t)\|f''\|_{[t,b],\infty}\right] \\ &\le \frac{1}{2}\cdot(b-t)(t-a)\|f''\|_{[a,b],\infty}, \quad t\in[a,b],\end{aligned} \tag{7.7}$$

where $f'' \in L_\infty[a,b]$. The constant $\frac{1}{2}$ is best possible in both inequalities. The function $f(t) = \frac{1}{2}\left(t-\frac{a+b}{2}\right)^2$ produces an equality in (7.7) for $t = \frac{a+b}{2}$.

If we assume that $\alpha = p$, $\beta = q$ in (7.2), then we also have:

$$|\Phi_f(t)| \le \frac{(b-t)(t-a)}{(q+1)^{1/q}(b-a)} \left[(t-a)^{1/q} \|f''\|_{[a,t],p} + (b-t)^{1/q} \|f''\|_{[t,b],p} \right] \le \frac{(b-t)(t-a)}{(q+1)^{1/q}(b-a)^{1/p}} \|f''\|_{[a,b],p}, \quad t \in [a,b], \tag{7.8}$$

for $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, $f'' \in L_p[a,b]$, since, by Hölder's inequality, we have

$$\begin{aligned} &(t-a)^{1/q} \|f''\|_{[a,t],p} + (b-t)^{1/q} \|f''\|_{[t,b],p} \\ &\le \left[(t-a)^{q/q} + (b-t)^{q/q} \right]^{1/q} \left[\|f''\|^p_{[a,t],p} + \|f''\|^p_{[t,b],p} \right]^{1/p} \\ &= (b-a)^{1/p} \|f''\|_{[a,b],p}. \end{aligned}$$

In the case that $p = q = 2$, we get the following inequality for the Euclidean norm $\|f''\|_{[a,b],2}$:

$$|\Phi_f(t)| \le \frac{\sqrt{3}}{3} \cdot \frac{(b-t)(t-a)}{b-a} \left[\sqrt{t-a} \|f''\|_{[a,t],2} + \sqrt{b-t} \|f''\|_{[t,b],2} \right] \le \frac{\sqrt{3}}{3} \cdot \frac{(b-t)(t-a)}{b-a} \|f''\|_{[a,b],2}, \quad t \in [a,b]. \tag{7.9}$$

It is an open question whether or not the constant $\frac{\sqrt{3}}{3}$ is best possible in (7.9).

Finally, from (7.2) we also have:

$$|\Phi_f(t)| \le \frac{(b-t)(t-a)}{b-a} \|f''\|_{[a,b],1} \le \frac{1}{4}(b-a) \|f''\|_{[a,b],1} \tag{7.10}$$

for any $t \in [a,b]$.

8. Applications for Weighted Means

For a function $f : [a,b] \to \mathbb{R}$, $\mathbf{x} = (x_1, \ldots, x_n) \in [a,b]^n$ and $\mathbf{p} = (p_1, \ldots, p_n)$ a probability sequence, i.e., $p_i \ge 0$, $i \in \{1, \ldots, n\}$ and $\sum_{i=1}^n p_i = 1$, we define the mean:

$$M_f(\mathbf{p}; \mathbf{x}) := \sum_{i=1}^n p_i f(x_i). \tag{8.1}$$

If $f(t) = t$, $t \in [a,b]$, then

$$M_f(\mathbf{p}; \mathbf{x}) = A(\mathbf{p}; \mathbf{x}) = \sum_{i=1}^n p_i x_i,$$

which is the *arithmetic mean* of $\mathbf{x}$ with the weights $\mathbf{p}$.

The main aim of the present section is to provide sharp bounds for the error in approximating $\mathcal{M}_f(\mathbf{p}; \mathbf{x})$ in terms of the simpler quantity

$$f(a) \cdot \frac{b - A(\mathbf{p}; \mathbf{x})}{b-a} + f(b) \cdot \frac{A(\mathbf{p}; \mathbf{x}) - a}{b-a}. \tag{8.2}$$

The following proposition contains some results of this type.

PROPOSITION 8.1 (Dragomir, 2008 [**53**]). *Let $f : [a,b] \to \mathbb{R}$ be a bounded function on $[a,b]$, $\mathbf{x} = (x_1, \ldots, x_n) \in [a,b]^n$ and $\mathbf{p}$ a probability sequence. Define the error functional $\mathcal{E}_f(\mathbf{p}; \mathbf{x})$ by:*

$$\mathcal{E}_f(\mathbf{p}; \mathbf{x}) := f(a) \cdot \frac{b - A(\mathbf{p}; \mathbf{x})}{b-a} + f(b) \cdot \frac{A(\mathbf{p}; \mathbf{x}) - a}{b-a} - \mathcal{M}_f(\mathbf{p}; \mathbf{x}). \tag{8.3}$$

(i) *If* $-\infty < m \leq f(t) \leq M < \infty$ *for any* $t \in [a,b]$, *then*

$$|\mathcal{E}_f(\mathbf{p};\mathbf{x})| \leq M - m. \tag{8.4}$$

The inequality is sharp.

(ii) *If* $f : [a,b] \to \mathbb{R}$ *is of bounded variation on* $[a,b]$, *then*

$$|\mathcal{E}_f(\mathbf{p};\mathbf{x})| \leq \left[\frac{1}{2} + \sum_{i=1}^{n} p_i \left|\frac{x_i - \frac{a+b}{2}}{b-a}\right|\right] \bigvee_a^b (f). \tag{8.5}$$

The constant $\frac{1}{2}$ *is best possible in (8.5).*

(iii) *If* $f : [a,b] \to \mathbb{R}$ *is* $L-$*Lipschitzian on* $[a,b]$, *then*

$$\begin{aligned} |\mathcal{E}_f(\mathbf{p};\mathbf{x})| &\leq \frac{2L}{b-a} \sum_{i=1}^{n} p_i (b - x_i)(x_i - a) \\ &\leq \frac{2L}{b-a} [b - A(\mathbf{p};\mathbf{x})][A(\mathbf{p};\mathbf{x}) - a] \leq \frac{1}{2} L (b-a). \end{aligned} \tag{8.6}$$

All the inequalities in (8.6) are sharp.

PROOF. Let us prove only the inequality (8.6). The other inequalities follow likewise. Applying the inequality (3.6) for $t = x_i$, $i \in \{1, \ldots, n\}$, we have

$$\left| f(x_i) - \frac{f(a)(b - x_i) + f(b)(x_i - a)}{b-a} \right| \leq \frac{2L}{b-a} (b - x_i)(x_i - a), \tag{8.7}$$

for any $i \in \{1, \ldots, n\}$. Multiplying (8.7) with p_i, summing over i from 1 to n and utilising the generalised triangle inequality $\sum_{i=1}^{n} |\alpha_i| \geq \left|\sum_{i=1}^{n} \alpha_i\right|$, we deduce the first inequality in (8.6).

Further, we use the following Čebyšev inequality:

$$\sum_{i=1}^{n} p_i \alpha_i \beta_i \leq \sum_{i=1}^{n} p_i \alpha_i \sum_{i=1}^{n} p_i \beta_i, \tag{8.8}$$

provided that $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$ and $(\alpha_i)_{i=\overline{1,n}}$, $(\beta_i)_{i=\overline{1,n}}$ are *asynchronous*, i.e.,

$$(\alpha_i - \alpha_j)(\beta_i - \beta_j) \leq 0 \quad \text{for any } i, j \in \{1, \ldots, n\}.$$

Then we have from (8.8)

$$\begin{aligned} \sum_{i=1}^{n} p_i (b - x_i)(x_i - a) &\leq \sum_{i=1}^{n} p_i (b - x_i) \sum_{i=1}^{n} p_i (x_i - a) \\ &= [b - A(\mathbf{p};\mathbf{x})][A(\mathbf{p};\mathbf{x}) - a] \end{aligned}$$

and the second inequality in (8.6) is proved. The last part is obvious.

The sharpness of the inequality follows from the case $n = 1$. The details are omitted. ∎

If $f : [a,b] \to \mathbb{R}$ is a convex function, then the following result is known in the literature as the (discrete) *Lah-Ribarić inequality*:

$$\sum_{i=1}^{n} p_i f(x_i) \leq \frac{1}{b-a} \{f(a)[b - A(\mathbf{p};\mathbf{x})] + f(b)[A(\mathbf{p};\mathbf{x}) - a]\}. \tag{8.9}$$

For a generalisation to positive linear functional that incorporates both the original Lah-Ribarić integral inequality and the discrete version of it due to Beesack and Pečaric [**8**], see [**113**, p. 98].

In terms of the error functional $\mathcal{E}_f(\mathbf{p};\mathbf{x})$, we then have $\mathcal{E}_f(\mathbf{p};\mathbf{x}) \geq 0,$ when f is convex and $\mathbf{p},\mathbf{x}$ are as above. Now, on utilising Theorem 2.2, we can state the following reverse of the Lah-Ribarić inequality (8.9).

PROPOSITION 8.2 (Dragomir, 2008 [**53**]). *If $f:[a,b]\to\mathbb{R}$ is convex on $[a,b]$ and the lateral derivatives $f'_-(b)$, $f'_+(a)$ are finite, then*

$$(0\leq)\,\mathcal{E}_f(\mathbf{p};\mathbf{x}) \leq \frac{f'_-(b)-f'_+(a)}{b-a}\sum_{i=1}^{n}p_i(b-x_i)(x_i-a) \leq \frac{f'_-(b)-f'_+(a)}{b-a}\left[b-A(\mathbf{p};\mathbf{x})\right]\left[A(\mathbf{p};\mathbf{x})-a\right] \leq \frac{1}{4}(b-a)\left[f'_-(b)-f'_+(a)\right]. \tag{8.10}$$

The inequalities are sharp and $\frac{1}{4}$ is best possible.

The following results in terms of the derivative of a function f can be stated as well.

PROPOSITION 8.3 (Dragomir, 2008 [**53**]). *Assume that $f:[a,b]\to\mathbb{R}$ is absolutely continuous on $[a,b]$.*

(i) *If f' is of bounded variation on $[a,b]$, then*

$$|\mathcal{E}_f(\mathbf{p};\mathbf{x})| \leq \frac{1}{b-a}\bigvee_a^b(f')\sum_{i=1}^{n}p_i(b-x_i)(x_i-a) \leq \frac{1}{b-a}\bigvee_a^b(f')\left[A(\mathbf{p};\mathbf{x})-a\right]\left[b-A(\mathbf{p};\mathbf{x})\right] \leq \frac{1}{4}(b-a)\bigvee_a^b(f'). \tag{8.11}$$

All inequalities in (8.11) are sharp. The constant $\frac{1}{4}$ is best possible.

(ii) *If f' is $K-$Lipschitzian on $[a,b]$ $(K>0)$, then*

$$|\mathcal{E}_f(\mathbf{p};\mathbf{x})| \leq \frac{1}{2}K\sum_{i=1}^{n}p_i(b-x_i)(x_i-a) \leq \frac{1}{2}K\left[b-A(\mathbf{p};\mathbf{x})\right]\left[A(\mathbf{p};\mathbf{x})-a\right] \leq \frac{1}{8}(b-a)^2K. \tag{8.12}$$

The constants $\frac{1}{2}$ and $\frac{1}{8}$ are best possible.

The proof is obvious by Theorem 5.2 and Theorem 6.1 and the details are omitted.

The above results can be useful in providing various inequalities between means. For instance, if we denote by $G(\mathbf{p},\mathbf{x})$ the *geometric mean* $\prod_{i=1}^{n}x_i^{p_i}$, then for the convex function $f(t)=-\ln t$, we have for $0<m\leq x_i\leq M<\infty$, $i\in\{1,\dots,n\}$ that:

$$\mathcal{E}_f(\mathbf{p};\mathbf{x}) = \ln G(\mathbf{p};\mathbf{x}) - \ln\left[m^{\frac{M-A(\mathbf{p};\mathbf{x})}{M-m}}\cdot M^{\frac{A(\mathbf{p};\mathbf{x})-m}{M-m}}\right] = \ln\left[\frac{G(\mathbf{p};\mathbf{x})}{m^{\frac{M-A(\mathbf{p};\mathbf{x})}{M-m}}\cdot M^{\frac{A(\mathbf{p};\mathbf{x})-m}{M-m}}}\right],$$

$$\bigvee_m^M(f) = \ln\left(\frac{M}{m}\right),$$

f is $L-$Lipschitzian with the constant $L=\|f'\|_{\infty,[m,M]}=\frac{1}{m}$ and

$$\frac{f'(M)-f'(m)}{M-m}=\frac{1}{mM},\qquad \bigvee_m^M(f')=\frac{M-m}{mM}.$$

Also, f' is $K-$Lipschitzian with the constant $K = \|f''\|_{\infty,[m,M]} = \frac{1}{m^2}$.

Applying Proposition 8.1, we get

$$0 \leq \ln \left[\frac{G(\mathbf{p};\mathbf{x})}{m^{\frac{M-A(\mathbf{p};\mathbf{x})}{M-m}} \cdot M^{\frac{A(\mathbf{p};\mathbf{x})-m}{M-m}}} \right] \leq \left[\frac{1}{2} + \sum_{i=1}^{n} p_i \left| \frac{x_i - \frac{m+M}{2}}{M-m} \right| \right] \ln \left(\frac{M}{m} \right),$$

while from Propositions 8.2 – 8.3 we get

$$\begin{aligned} 0 &\leq \ln \left[\frac{G(\mathbf{p};\mathbf{x})}{m^{\frac{M-A(\mathbf{p};\mathbf{x})}{M-m}} \cdot M^{\frac{A(\mathbf{p};\mathbf{x})-m}{M-m}}} \right] \\ &\leq \min \left\{ \frac{2}{m(M-m)}, \frac{1}{mM}, \frac{1}{2m^2} \right\} \cdot \sum_{i=1}^{n} p_i (M - x_i)(x_i - m) \\ &\leq \min \left\{ \frac{2}{m(M-m)}, \frac{1}{mM}, \frac{1}{2m^2} \right\} \cdot [M - A(\mathbf{p};\mathbf{x})][A(\mathbf{p};\mathbf{x}) - m] \\ &\leq \min \left\{ \frac{M-m}{2m}, \frac{(M-m)^2}{4mM}, \frac{(M-m)^2}{8m^2} \right\}. \end{aligned}$$

REMARK 8.1. All the results in this section can be stated for positive linear functionals defined on linear spaces of functions. Applications for Lebesgue integrals in the general setting of measurable spaces can be provided as well. However, for the sake of brevity, we do not state them here.

9. APPLICATIONS FOR f-DIVERGENCES

Now, for $0 < r < 1 < R < \infty$ we consider the expression

$$\frac{1}{R-r} [(R-1) f(r) + (1-r) f(R)]$$

and are interested to compare it with the f-divergence $I_f(\mathbf{p},\mathbf{q})$ which can be extended for larger classes than convex functions with the same definition and the same conventions as in the case of convex functions outlined at the begining of the book.

PROPOSITION 9.1 (Dragomir, 2008 [**53**]). *Let $f : [r,R] \to \mathbb{R}$ be a bounded function on the interval $[r,R]$ with $0 < r < 1 < R < \infty$. Assume that $\mathbf{p},\mathbf{q} \in \mathbb{P}^n$ are such that*

$$r \leq \frac{p_i}{q_i} \leq R \text{ for each } i \in \{1,...,n\} \tag{9.1}$$

and define the error functional

$$\delta_f(\mathbf{p},\mathbf{q};r,R) := \frac{1}{R-r} [(R-1) f(r) + (1-r) f(R)] - I_f(\mathbf{p},\mathbf{q}).$$

(i) *If $-\infty < m \leq f(t) \leq M < \infty$ for any $t \in [r,R]$, then*

$$|\delta_f(\mathbf{p},\mathbf{q};r,R)| \leq M - m. \tag{9.2}$$

The inequality is sharp.

(ii) *If $f : [r,R] \to \mathbb{R}$ is of bounded variation on $[r,R]$, then*

$$|\delta_f(\mathbf{p},\mathbf{q};r,R)| \leq \left[\frac{1}{2} + \frac{1}{R-r} \sum_{i=1}^{n} \left| p_i - \frac{r+R}{2} \cdot q_i \right| \right] \bigvee_r^R (f). \tag{9.3}$$

The constant $\frac{1}{2}$ is best possible in (8.5).

(iii) *If* $f:[r,R]\to\mathbb{R}$ *is* $L-$*Lipschitzian on* $[r,R]$, *then*

$$|\delta_f(\mathbf{p},\mathbf{q};r,R)| \le \frac{2L}{R-r}\left[(R-1)(1-r)-\chi^2(p,q)\right] \le \frac{2L}{R-r}(R-1)(1-r) \le \frac{1}{2}L(R-r), \tag{9.4}$$

where the K. Pearson χ^2*-divergence is obtained from the general case for the convex function* $f(t)=(1-t)^2$, $t\in\mathbb{R}$ *and given by the equivalent expressions:*

$$\chi^2(p,q) := \sum_{j=1}^{n} q_j\left(\frac{p_j}{q_j}-1\right)^2 = \sum_{j=1}^{n}\frac{(p_j-q_j)^2}{q_j} = \sum_{j=1}^{n}\frac{p_j^2}{q_j}-1. \tag{9.5}$$

PROOF. The proof follows in a similar manner with the one from Proposition 8.1 on choosing $a=r, b=R, p_i=q_i$ and $x_i=\frac{p_i}{q_i}$ with $i\in\{1,...,n\}$ and the details are omitted. ∎

In the case of convex functions we have

PROPOSITION 9.2 (Dragomir, 2008 [**53**]). *If* $f:[r,R]\to\mathbb{R}$ *is convex on* $[r,R]$ *and the lateral derivatives* $f'_-(R)$, $f'_+(r)$ *are finite, then*

$$(0\le)\,\delta_f(\mathbf{p},\mathbf{q};r,R) \le \frac{f'_-(R)-f'_+(r)}{R-r}\left[(R-1)(1-r)-\chi^2(p,q)\right] \le \frac{f'_-(R)-f'_+(r)}{R-r}(R-1)(1-r) \le \frac{1}{4}(R-r)\left[f'_-(R)-f'_+(r)\right], \tag{9.6}$$

provided that $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ *are such that (9.1) holds.*

The inequalities are sharp and $\frac{1}{4}$ *is best possible.*

We notice that the result from Proposition 9.2 has been firstly obtained by the author in the paper [**36**].

Finally, we can state:

PROPOSITION 9.3 (Dragomir, 2008 [**53**]). *Assume that* $f:[r,R]\to\mathbb{R}$ *is absolutely continuous on* $[r,R]$ *and that* $\mathbf{p},\mathbf{q}\in\mathbb{P}^n$ *are such that (9.1) holds.*

(i) *If* f' *is of bounded variation on* $[r,R]$, *then*

$$|\delta_f(\mathbf{p},\mathbf{q};r,R)| \le \frac{1}{R-r}\bigvee_r^R(f')\left[(R-1)(1-r)-\chi^2(p,q)\right] \le \frac{1}{R-r}(R-1)(1-r)\bigvee_r^R(f') \le \frac{1}{4}(R-r)\bigvee_r^R(f'). \tag{9.7}$$

All inequalities in (9.7) are sharp. The constant $\frac{1}{4}$ *is best possible.*

(ii) *If* f' *is* $K-$*Lipschitzian on* $[r,R]$ $(K>0)$, *then*

$$|\delta_f(\mathbf{p},\mathbf{q};r,R)| \le \frac{1}{2}K\left[(R-1)(1-r)-\chi^2(p,q)\right] \le \frac{1}{2}K(R-1)(1-r) \le \frac{1}{8}(R-r)^2K. \tag{9.8}$$

The constants $\frac{1}{2}$ *and* $\frac{1}{8}$ *are best possible.*

CHAPTER 15

Approximation of Kullback-Leibler Distance

1. Some Inequalities for the Logarithmic Mapping

The following theorem is well known in the literature as Taylor's formula or Taylor's theorem with the integral remainder.

THEOREM 1.1. *Let $I \subset \mathbb{R}$ be a closed interval, $a \in I$ and let n be a positive integer. If $f : I \longrightarrow \mathbb{R}$ is such that $f^{(n)}$ is absolutely continuous on I, then for each $x \in I$*

$$f(x) = T_n(f; a, x) + R_n(f; a, x) \tag{1.1}$$

where $T_n(f; a, x)$ is Taylor's polynomial, i.e.,

$$T_n(f; a, x) := \sum_{k=0}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a).$$

(Note that $f^{(0)} := f$ and $0! := 1$), and the remainder is given by

$$R_n(f; a, x) := \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t)\, dt.$$

A simple proof of this theorem can be achieved by mathematical induction using the integration by parts formula.

The following corollary concerning the estimation of the remainder is useful when we want to approximate real functions by their Taylor expansions.

COROLLARY 1.2. *With the above assumptions, we have the estimation*

$$|R_n(f; a, x)| \leq \frac{|x-a|^n}{n!} \left| \int_a^x \left| f^{(n+1)}(t) \right| dt \right| \tag{1.2}$$

or

$$|R_n(f; a, x)| \leq \frac{1}{n!} \frac{|x-a|^{n+\frac{1}{\beta}}}{(n\beta+1)^{\frac{1}{\beta}}} \left| \int_a^x \left| f^{(n+1)}(t) \right|^{\alpha} dt \right|^{\frac{1}{\alpha}} \tag{1.3}$$

where $\alpha > 1$ and $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, and the estimation:

$$|R_n(f; a, x)| \leq \frac{|x-a|^{n+1}}{(n+1)!} \max\left\{ \left| f^{(n+1)}(t) \right|,\ t \in [a, x] \text{ or } [x, a] \right\} \tag{1.4}$$

respectively.

PROOF. The inequalities (1.2) and (1.4) are obvious.
Using Hölder's integral inequality, we have that

$$\left|\int_a^x (x-t)^n f^{(n+1)}(t)\,dt\right| \le \left|\int_a^x \left|f^{(n+1)}(t)\right|^\alpha dt\right|^{\frac{1}{\alpha}} \left|\int_a^x |x-t|^{n\beta} dt\right|^{\frac{1}{\beta}}$$
$$= \left[\frac{|x-a|^{n\beta+1}}{n\beta+1}\right]^{\frac{1}{\beta}} \left|\int_a^x \left|f^{(n+1)}(t)\right|^\alpha dt\right|^{\frac{1}{\alpha}}$$

and the inequality (1.3) is also proved. ∎

The following result for the logarithmic mapping holds.

COROLLARY 1.3. *Let $a,b>0$, then we have the equality:*

$$\ln b - \ln a - \frac{b-a}{a} + \sum_{k=2}^{n} \frac{(-1)^k (b-a)^k}{ka^k} = (-1)^n \int_a^b \frac{(b-t)^n}{t^{n+1}} dt. \tag{1.5}$$

PROOF. Consider the mapping $f:(0,\infty)\longrightarrow\mathbb{R}$, $f(x)=\ln x$, then,

$$f^{(n)}(x) = \frac{(-1)^{n-1}(n-1)!}{x^n},\ n\ge 1,\ x>0,$$
$$T_n(f;a,x) = \ln a + \sum_{k=1}^{n} \frac{(-1)^{k-1}(x-a)^k}{ka^k},\ a>0$$

and

$$R_n(f;a,x) = (-1)^n \int_a^x \frac{(x-t)^n}{t^{n+1}} dt.$$

Now, using (1.1), we have the equality,

$$\ln x = \ln a + \sum_{k=1}^{n} \frac{(-1)^{k-1}(x-a)^k}{ka^k} + (-1)^n \int_a^x \frac{(x-t)^n}{t^{n+1}} dt,$$

i.e.,

$$\ln x - \ln a + \sum_{k=1}^{n} \frac{(-1)^k (x-a)^k}{ka^k} = (-1)^n \int_a^x \frac{(x-t)^n}{t^{n+1}} dt, \quad x,a>0.$$

Choosing in the last equality $x=b$, we get (1.5). ∎

The following inequality for logarithms holds.

COROLLARY 1.4. *For all $a,b>0$, we have the inequality:*

$$\left|\ln b - \ln a - \frac{b-a}{a} + \sum_{k=2}^{n} \frac{(-1)^k (b-a)^k}{ka^k}\right| \tag{1.6}$$
$$\le \begin{cases} \frac{|b-a|^n |b^n-a^n|}{na^nb^n}; \\ \frac{|b-a|^{n+\frac{1}{\beta}}}{[(n+1)\alpha-1]^{\frac{1}{\alpha}}(n\beta+1)^{\frac{1}{\beta}}} \left[\frac{\left|b^{(n+1)\alpha-1}-a^{(n+1)\alpha-1}\right|}{b^{(n+1)\alpha-1}a^{(n+1)\alpha-1}}\right]^{\frac{1}{\alpha}}, \quad \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{|b-a|^{n+1}}{n+1}\left[\frac{1}{\min\{a,b\}}\right]^{n+1}. \end{cases}$$

The equality holds if and only if $a=b$.

PROOF. We use Corollary 1.2 for mapping $f : (0,\infty) \to \mathbb{R}$, $f(x) = \ln x$ for which we have,

$$\int_a^b \left|f^{(n+1)}(t)\right| dt = n! \int_a^b \frac{dt}{t^{n+1}} = n! \left[\frac{t^{-n+1-1}}{-n+1-1}\bigg|_a^b \right]$$
$$= \frac{n!}{n}\left[\frac{1}{a^n} - \frac{1}{b^n}\right] = \frac{n!}{n} \cdot \frac{b^n - a^n}{a^n b^n}.$$

Also,

$$\int_a^b \left|f^{(n+1)}(t)\right|^\alpha dt = (n!)^\alpha \int_a^b \frac{dt}{t^{\alpha(n+1)}} = \frac{(n!)^\alpha}{(n+1)\alpha - 1} \cdot \frac{b^{(n+1)\alpha-1} - a^{(n+1)\alpha-1}}{b^{(n+1)\alpha-1} \cdot a^{(n+1)\alpha-1}}$$

and

$$\max\left\{\left|f^{(n+1)}(t)\right|,\ t\in[a,b] \text{ or } t\in[b,a]\right\}$$
$$= \max\left\{n!\frac{1}{t^{n+1}},\ t\in[a,b] \text{ or } t\in[b,a]\right\}$$
$$= n!\frac{1}{\min\{a^{n+1},b^{n+1}\}} = n!\left[\frac{1}{\min\{a,b\}}\right]^{n+1}.$$

The equality in (1.6) holds via the representation (1.5), the details are omitted. ∎

REMARK 1.1. By the concavity property of $\ln(\cdot)$ we have

$$\ln b - \ln a \le \frac{b-a}{a}$$

and then, if we choose $n = 1$ in (1.6), we get the following counterpart result.

$$0 \le \frac{b-a}{a} - \ln b + \ln a$$
$$\le \begin{cases} \frac{(b-a)^2}{ab}; \\ \frac{|b-a|^{1+\frac{1}{\beta}}\left|b^{2\alpha-1}-a^{2\alpha-1}\right|^{\frac{1}{\alpha}}}{(2\alpha-1)^{\frac{1}{\alpha}}(\beta+1)^{\frac{1}{\beta}}a^{2-\frac{1}{\alpha}}b^{2-\frac{1}{\alpha}}}, & \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{(b-a)^2}{2}\cdot\frac{1}{\min^2\{a,b\}}. \end{cases}$$

The equality holds in both inequalities simultaneously if and only if $a = b$.

REMARK 1.2. If we choose $n = 2$ in (1.6), we get,

$$\left|\ln b - \ln a - \frac{b-a}{a} + \frac{(b-a)^2}{2a^2}\right|$$
$$\le \begin{cases} \frac{(b-a)^3}{a^2b^2}\cdot\frac{a+b}{2}; \\ \frac{|b-a|^{2+\frac{1}{\beta}}\left|b^{3\alpha-1}-a^{3\alpha-1}\right|^{\frac{1}{\alpha}}}{(3\alpha-1)^{\frac{1}{\alpha}}(2\beta+1)^{\frac{1}{\beta}}a^{3-\frac{1}{\alpha}}b^{3-\frac{1}{\alpha}}}, & \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{(b-a)^3}{3}\cdot\frac{1}{\min^3\{a,b\}}. \end{cases}$$

The equality holds in both inequalities simultaneously if and only if $a = b$.

2. INEQUALITIES FOR RELATIVE ENTROPY

Let X and Y be two random variables having the probability $p_i, q_i, i \in \{1, ..., m\}$, then we have the following representation of relative entropy [**63**].

THEOREM 2.1 (Dragomir & Gluščević, 2001 [**63**]). *With the above assumptions of X and Y, we have,*

$$D_{KL}(p,q) = \sum_{k=2}^{n} \sum_{i=1}^{m} \frac{(p_i - q_i)^k}{k p_i^{k-1}} + (-1)^{n-1} \sum_{i=1}^{m} p_i \left(\int_{p_i}^{q_i} \frac{(q_i - t)^n}{t^{n+1}} dt \right) \tag{2.1}$$

or

$$D_{KL}(p,q) = \sum_{k=1}^{n} \sum_{i=1}^{m} p_i \left(\frac{(-1)^{k-1} (p_i - q_i)^k}{k q_i^k} \right) + (-1)^{n} \sum_{i=1}^{m} p_i \left(\int_{q_i}^{p_i} \frac{(p_i - t)^n}{t^{n+1}} dt \right). \tag{2.2}$$

PROOF. Choose $a = p_i, b = q_i, i \in \{1, ..., m\}$ in (1.5) to get,

$$\ln q_i - \ln p_i - \frac{q_i - p_i}{p_i} + \sum_{k=2}^{n} \frac{(-1)^k (q_i - p_i)^k}{k p_i^k} = (-1)^n \int_{p_i}^{q_i} \frac{(q_i - t)^n}{t^{n+1}} dt. \tag{2.3}$$

Multiply (2.3) by p_i and sum over i to obtain,

$$- D_{KL}(p,q) - \sum_{i=1}^{m} (q_i - p_i) + \sum_{i=1}^{m} \left(\sum_{k=2}^{n} \frac{(-1)^k (q_i - p_i)^k}{k p_i^{k-1}} \right) = (-1)^n \sum_{i=1}^{m} p_i \left(\int_{p_i}^{q_i} \frac{(q_i - t)^n}{t^{n+1}} dt \right). \tag{2.4}$$

However,

$$\sum_{i=1}^{m} (q_i - p_i) = 0,$$

therefore, by (2.4) we get (2.1) .
To prove the second equality, choose $b = p_i, a = q_i, i \in \{1, ..., m\}$ in (1.5) to get,

$$\ln p_i - \ln q_i - \frac{p_i - q_i}{q_i} + \sum_{k=2}^{n} \frac{(-1)^k (p_i - q_i)^k}{k q_i^k} = (-1)^n \int_{q_i}^{p_i} \frac{(p_i - t)^n}{t^{n+1}} dt. \tag{2.5}$$

Multiply (2.5) by p_i and sum over i to get,

$$D_{KL}(p,q)=\sum_{i=1}^{m}p_i\left(\sum_{k=1}^{n}\frac{(-1)^{k-1}(p_i-q_i)^k}{kq_i^k}\right)+(-1)^n\sum_{i=1}^{m}p_i\left(\int_{q_i}^{p_i}\frac{(p_i-t)^n}{t^{n+1}}dt\right)$$

from which we deduce (2.2). ∎

Using Corollary 1.4, we can give the following result containing an approximation of the relative entropy [**63**].

THEOREM 2.2 (Dragomir & Glušćević, 2001 [**63**]). *With the above assumption over X and Y, we have,*

$$\left|D_{KL}(p,q)-\sum_{k=2}^{n}\sum_{i=1}^{m}\frac{(p_i-q_i)^k}{kp_i^{k-1}}\right| \le M:=\begin{cases}\frac{1}{n}\sum_{i=1}^{m}\frac{|q_i-p_i|^n\left|q_i^n-p_i^n\right|}{p_i^{n-1}q_i^n};\\ \frac{1}{[(n+1)\alpha-1]^{\frac{1}{\alpha}}(n\beta+1)^{\frac{1}{\beta}}}\sum_{i=1}^{m}p_i\,|q_i-p_i|^{n+\frac{1}{\beta}}\\ \qquad\times\left[\frac{\left|q_i^{(n+1)\alpha-1}-p_i^{(n+1)\alpha-1}\right|}{q_i^{(n+1)\alpha-1}p_i^{(n+1)\alpha-1}}\right]^{\frac{1}{\alpha}};\\ \frac{1}{n+1}\sum_{i=1}^{m}p_i\,|q_i-p_i|^{n+1}\\ \qquad\times\left[\frac{1}{\min\{p_i,q_i\}}\right]^{n+1};\end{cases}$$

and

$$\left|D_{KL}(p,q)-\sum_{i=1}^{m}p_i\left(\sum_{k=1}^{n}\frac{(-1)^{k-1}(p_i-q_i)^k}{kq_i^k}\right)\right|\le M.$$

PROOF. Proof of the first inequality is obvious by Corollary 1.4, choosing $a=p_i$, $b=q_i$, multiplying by q_i and summing over $i\in\{1,...,m\}$. Proof for second inequality is obvious by Corollary 1.4, choosing $b=p_i$, $a=q_i$, multiplying by q_i and summing over $i\in\{1,...,m\}$. ∎

COROLLARY 2.3 (Dragomir & Glušćević, 2001 [**63**]). *Under the assumptions of Theorem 2.2 for $n=1$, we have,*

$$D_{KL}(p,q)\le M_1 \tag{2.6}$$

where

$$M_1:=\begin{cases}\sum_{i=1}^{m}\frac{(q_i-p_i)^2}{q_i}=D_{\chi^2}(q,p);\\ \frac{1}{(2\alpha-1)^{\frac{1}{\alpha}}(\beta+1)^{\frac{1}{\beta}}}\sum_{i=1}^{m}p_i\,|q_i-p_i|^{1+\frac{1}{\beta}}\times\frac{\left|q_i^{2\alpha-1}-p_i^{2\alpha-1}\right|^{\frac{1}{\alpha}}}{q_i^{2-\frac{1}{\alpha}}p_i^{2-\frac{1}{\alpha}}};\\ \frac{1}{2}\sum_{i=1}^{m}p_i\,(q_i-p_i)^2\times\frac{1}{\min^2\{p_i,q_i\}};\end{cases}$$

and

$$0\le D_{\chi^2}(q,p)-D_{KL}(p,q)\le M_1, \tag{2.7}$$

where $\alpha>1$ *and* $\frac{1}{\alpha}+\frac{1}{\beta}=1$.

REMARK 2.1. The first inequality in (2.6) is equivalent to (see also [**78**]):

$$D_{KL}(p,q) \leq D_{\chi^2}(q,p) - 1.$$

We introduce the notation $M_{\alpha,\beta}$ for the second term in M_1 and apply Hölder's inequality. We can then write,

$$\begin{aligned} M_{\alpha,\beta} &:= \sum_{i=1}^{m} p_i |q_i - p_i|^{\frac{\beta+1}{\beta}} \times \frac{\left|q_i^{2\alpha-1} - p_i^{2\alpha-1}\right|^{\frac{1}{\alpha}}}{q_i^{2-\frac{1}{\alpha}} p_i^{2-\frac{1}{\alpha}}} \\ &\leq \left(\sum_{i=1}^{m} p_i |q_i - p_i|^{\beta+1}\right)^{\frac{1}{\beta}} \left(\sum_{i=1}^{m} \frac{\left|q_i^{2\alpha-1} - p_i^{2\alpha-1}\right|}{q_i^{2\alpha-1} p_i^{2\alpha-2}}\right)^{\frac{1}{\alpha}} \\ &:= \tilde{M}_{\alpha,\beta} \end{aligned}$$

and from the second inequality of (2.6) we obtain,

$$D_{KL}(p,q) \leq \frac{1}{(2\alpha-1)^{\frac{1}{\alpha}} (\beta+1)^{\frac{1}{\beta}}} \tilde{M}_{\alpha,\beta}.$$

For $\alpha = \beta = 2$, we get the particular inequality,

$$D_{KL}(p,q) \leq \frac{1}{3} \left(\sum_{i=1}^{m} p_i |q_i - p_i|^3\right)^{\frac{1}{2}} \left(\sum_{i=1}^{m} \frac{|q_i^3 - p_i^3|}{q_i^3 p_i^2}\right)^{\frac{1}{2}}.$$

If we assume that

$$\inf_{x\in\Gamma} [\min\{p_i, q_i\}] = \delta > 0 \tag{2.8}$$

then from the third inequality of (2.6) we have,

$$D_{KL}(p,q) \leq \frac{1}{2\delta^2} \sum_{i=1}^{m} p_i (q_i - p_i)^2.$$

REMARK 2.2. Since

$$\sum_{i=1}^{m} \frac{p_i}{q_i} (p_i - q_i) = D_{\chi^2}(q,p),$$

then the first inequality in (2.7) is obvious.
Using Hölder's inequality, from the second inequality in (2.7) we get,

$$D_{\chi^2}(q,p) - D_{KL}(p,q) \leq \frac{1}{(2\alpha-1)^{\frac{1}{\alpha}} (\beta+1)^{\frac{1}{\beta}}} \tilde{M}_{\alpha,\beta}.$$

If, as above, we assume that (2.8) holds, then from the third inequality in (2.7) we have,

$$D_{\chi^2}(q,p) - D_{KL}(p,q) \leq \frac{1}{2\delta^2} \sum_{i=1}^{m} p_i (q_i - p_i)^2.$$

3. INEQUALITIES FOR THE ENTROPY MAPPING

Let X be a random variable having the probability mass function p_i, $i \in \{1, ..., m\}$. Consider the *entropy mapping,*

$$H(X) = \sum_{i=1}^{m} p_i \left(\ln \frac{1}{p_i}\right).$$

We have the following representation of $H(X)$ [**63**].

THEOREM 3.1 (Dragomir & Gluščević, 2001 [**63**]). *With the above assumption for X, we have,*

$$H(X) = \ln m - \sum_{k=2}^{n} \sum_{i=1}^{m} \frac{(mp_i - 1)^k}{km^k p^{k-1}(x)} + \frac{(-1)^n}{m^n} \sum_{i=1}^{m} p_i \left(\int_{p_i}^{\frac{1}{m}} \frac{(1-mt)^n}{t^{n+1}} dt \right) \tag{3.1}$$

or

$$H(X) = \ln m + \sum_{k=1}^{n} \sum_{i=1}^{m} p_i \left(\frac{(-1)^k (mp_i - 1)^k}{k} \right) + (-1)^{n+1} \sum_{i=1}^{m} p_i \left(\int_{\frac{1}{m}}^{p_i} \frac{(p_i - t)^n}{t^{n+1}} dt \right). \tag{3.2}$$

PROOF. Put $q = u$ in (2.1), where u is the uniform distribution, i.e., $u_i = \frac{1}{m}$, then,

$$\begin{aligned} \ln m - H(X) &= \sum_{k=2}^{n} \sum_{i=1}^{m} \frac{\left(p_i - \frac{1}{m}\right)^k}{kp_i^{k-1}} \\ &\quad + (-1)^{n-1} \sum_{i=1}^{m} p_i \left(\int_{p_i}^{\frac{1}{m}} \frac{\left(\frac{1}{m} - t\right)^n}{t^{n+1}} dt \right) \\ &= \sum_{k=2}^{n} \sum_{i=1}^{m} \frac{(mp_i - 1)^k}{km^k p_i^{k-1}} \\ &\quad + \frac{(-1)^{n-1}}{m^n} \sum_{i=1}^{m} p_i \left(\int_{p_i}^{\frac{1}{m}} \frac{(1-mt)^n}{t^{n+1}} dt \right) \end{aligned}$$

from which results (3.1).
Put $q = u$ in (2.2) to get,

$$\begin{aligned} \ln m - H(X) &= \sum_{i=1}^{m} p_i \sum_{k=1}^{n} \frac{(-1)^{k-1} \left(p_i - \frac{1}{m}\right)^k}{k \frac{1}{m^k}} \\ &\quad + (-1)^n \sum_{i=1}^{m} p_i \left(\int_{\frac{1}{m}}^{p_i} \frac{(p_i - t)^n}{t^{n+1}} dt \right) \\ &= \sum_{k=1}^{n} \sum_{i=1}^{m} p_i \frac{(-1)^{k-1} (mp_i - 1)^k}{k} \\ &\quad + (-1)^n \sum_{i=1}^{m} p_i \left(\int_{\frac{1}{m}}^{p_i} \frac{(p_i - t)^n}{t^{n+1}} dt \right), \end{aligned}$$

from which results (3.2). ∎

Using Theorem 2.2, we can state the following result concerning the approximation of the entropy mapping [**63**].

THEOREM 3.2 (Dragomir & Gluščević, 2001 [**63**]). *With the above assumption for X, we have,*

$$\left|H(X)-\ln m+\sum_{k=2}^{n}\sum_{i=1}^{m}\frac{(mp_i-1)^k}{km^kp_i^{k-1}}\right| \le \mu := \begin{cases} \frac{1}{nm^n}\sum_{i=1}^{m}\frac{|1-mp_i|^n\left|1-[m]^np_i^n\right|}{p_i^{n-1}}; \\ \frac{1}{[(n+1)\alpha-1]^{\frac{1}{\alpha}}(n\beta+1)^{\frac{1}{\beta}}}\sum_{i=1}^{m}p_i\left|\frac{1}{m}-p_i\right|^{n+\frac{1}{\beta}} \times\left[\frac{\left|1-m^{(n+1)\alpha-1}p_i^{(n+1)\alpha-1}\right|}{p_i^{(n+1)\alpha-1}}\right]^{\frac{1}{\alpha}}; \\ \frac{1}{n+1}\sum_{i=1}^{m}p_i\left|\frac{1}{m}-p_i\right|^{n+1}\times\left[\frac{1}{\min\left\{p_i,\frac{1}{m}\right\}}\right]^{n+1}; \end{cases}$$

and

$$\left|H(X)-\ln m-\sum_{i=1}^{m}p_i\sum_{k=1}^{n}\frac{(-1)^k(mp_i-1)^k}{k}\right|\le\mu.$$

4. INEQUALITIES FOR MUTUAL INFORMATION

Let X and Y be random variables having the probability distributions $p_i, q_j, i\in\{1,...,m\}$, $j\in\{1,...,m\}$. Consider the *mutual information* [**21**],

$$I(X,Y)=\sum_{i=1}^{m}\sum_{i=1}^{m}p_{ij}\ln\frac{p_{ij}}{p_iq_j},$$

where $p_{ij}, i\in\{1,...,m\}, j\in\{1,...,m\}$ is the mutual probability distribution.

We have the following representation for $I(X,Y)$ [**63**].

THEOREM 4.1 (Dragomir & Gluščević, 2001 [**63**]). *With the above assumption for X and Y,*

$$I(X,Y)=\sum_{k=2}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{(p_{ij}-p_iq_j)^k}{kp_{ij}^{k-1}} +(-1)^{n-1}\sum_{j=1}^{m}\sum_{i=1}^{m}p_{ij}\left(\int_{p_{ij}}^{p_iq_j}\frac{(p_iq_j-t)^n}{t^{n+1}}dt\right)$$

or

$$I(X,Y)=\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}p_{ij}\left(\frac{(-1)^{k-1}(p_{ij}-p_iq_j)^k}{kp_i^kq_j^k}\right) +(-1)^n\sum_{i=1}^{m}\sum_{j=1}^{m}p_{ij}\left(\int_{p_iq_j}^{p_{ij}}\frac{(p_{ij}-t)^n}{t^{n+1}}dt\right).$$

The proof follows by Theorem 2.1, taking into account that

$$I(X,Y)=D(p_{ij}\,\|p_iq_j).$$

Finally, using Theorem 2.2, we can state the following estimation of the *mutual information.*

THEOREM 4.2 (Dragomir & Gluščević, 2001 [**63**]). *With the above assumption over X and Y, we have:*

$$\left| I(X,Y) - \sum_{k=2}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{(p_{ij}-p_iq_j)^k}{kp_{ij}^{k-1}}\right|$$

$$\leq \tilde{M} := \begin{cases} \frac{1}{n}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{|p_iq_j-p_{ij}|^n}{p_{ij}^{n-1}p_i^{n-1}q_j^{n-1}(y)} \times \left|p_i^nq_j^n - p_{ij}^n\right|; \\ \\ \frac{1}{[(n+1)\alpha-1]^{\frac{1}{\alpha}}(n\beta+1)^{\frac{1}{\beta}}}\sum_{i=1}^{m}\sum_{j=1}^{m}p_{ij}\times|p_iq_j-p_{ij}|^{n+\frac{1}{\beta}} \\ \qquad \times\left[\frac{\left|p_i^{(n+1)\alpha-1}q_j^{(n+1)\alpha-1}-p_{ij}^{(n+1)\alpha-1}\right|}{p_i^{(n+1)\alpha-1}q_j^{(n+1)\alpha-1}p_{ij}^{(n+1)\alpha-1}}\right]^{\frac{1}{\alpha}}; \\ \\ \frac{1}{n+1}\sum_{i=1}^{m}\sum_{j=1}^{m}p_{ij}\times|p_iq_j-p_{ij}|^{n+1} \\ \qquad \times\left[\frac{1}{\min\{p_{ij},p_iq_j\}}\right]^{n+1}; \end{cases}$$

and

$$\left| I(X,Y) - \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{m}p_{ij}\frac{(-1)^{k-1}(p_{ij}-p_iq_j)^k}{kp_i^kq_j^k}\right| \leq \tilde{M}.$$

CHAPTER 16

Approximation of f-Divergences Via Taylor Expansions

1. General Results

We start with the following result [**4**].

THEOREM 1.1 (Barnett et al., 2002 [**4**]). *Let $f:(0,\infty)\to\mathbb{R}$ be $n-$time differentiable and such that $f^{(n)}$ is absolutely continuous on $[r,R]$, where $0<r\leq 1\leq R<\infty$. Assume that the probability distributions p,q satisfy the condition,*

$$r\leq\frac{p_i}{q_i}\leq R,\quad i\in\{1,...,m\}. \tag{1.1}$$

We then have,

$$\left|I_f(p,q)-f(1)-\sum_{k=1}^{n}\frac{f^{(k)}(1)}{k!}D_{\chi^k}(p,q)\right| \tag{1.2}$$

$$\leq\begin{cases}\frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty D_{|\chi|^{n+1}}(p,q) & \textit{if}\quad f^{(n+1)}\in L_\infty[r,R];\\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha D_{|\chi|^{n+\frac{1}{\beta}}}(p,q) & \textit{if}\quad f^{(n+1)}\in L_\alpha[r,R],\\ & \qquad\alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1;\\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_1 D_{|\chi|^n}(p,q);\end{cases}$$

$$\leq\begin{cases}\frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty (R-r)^{n+1} & \textit{if}\quad f^{(n+1)}\in L_\infty[r,R];\\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha (R-r)^{n+\frac{1}{\beta}} & \textit{if}\quad f^{(n+1)}\in L_\alpha[r,R],\\ & \qquad\alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1;\\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_1 (R-r)^n;\end{cases}$$

where

$$D_{\chi^k}(p,q):=\sum_{i=1}^{m}\frac{(p_i-q_i)^k}{q_i^{k-1}},$$

$$D_{|\chi|^s}(p,q):=\sum_{i=1}^{m}\frac{|p_i-q_i|^s}{q_i^{s-1}}\quad(k\in\mathbb{N},\ s\geq 0)$$

and $\|\cdot\|_\alpha$ are the usual Lebesgue norms, i.e.,

$$\left\|f^{(n+1)}\right\|_\alpha:=\left(\int_r^R\left|f^{(n+1)}\right|^\alpha dt\right)^{\frac{1}{\alpha}},\quad\alpha\geq 1,$$

$$\left\|f^{(n+1)}\right\|_\infty:=ess\sup_{t\in[r,R]}\left|f^{(n+1)}(t)\right|.$$

PROOF. We start with a Taylor expansion with integral remainder:

$$f(z)=f(a)+\sum_{k=1}^{n}\frac{(z-a)^k}{k!}f^{(k)}(a)+\frac{1}{n!}\int_a^z (z-t)^n f^{(n+1)}(t)\,dt \tag{1.3}$$

for all $z,a\in(0,\infty)$.

Using the properties of the modulus, we have,

$$\begin{aligned}\left|f(z)-f(a)-\sum_{k=1}^{n}\frac{(z-a)^k}{k!}f^{(k)}(a)\right| &\le \frac{1}{n!}\left|\int_a^z |z-t|^n \left|f^{(n+1)}(t)\right| dt\right| \\ &:= M\left(f^{(n+1)};a,z\right).\end{aligned}\tag{1.4}$$

Now, assume that $a,z\in[r,R]$, then, obviously,

$$\begin{aligned}M\left(f^{(n+1)};a,z\right) &\le \sup_{t\in[r,R]}\left|f^{(n+1)}(t)\right|\frac{1}{n!}\left|\int_a^z |z-t|^n dt\right| \\ &= \frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty |z-a|^{n+1},\end{aligned}\tag{1.5}$$

for all $a,z\in[r,R]$.

Also, by use of Hölder's integral inequality, we have:

$$\begin{aligned}&M\left(f^{(n+1)};a,z\right) \\ &\le \frac{1}{n!}\left|\int_a^z |x-t|^{n\beta}dt\right|^{\frac{1}{\beta}}\left[\int_a^z \left|f^{(n+1)}(t)\right|^\alpha dt\right]^{\frac{1}{\alpha}} \\ &\le \frac{1}{n!}\left\|f^{(n+1)}\right\|_\alpha\left[\frac{|z-a|^{n\beta+1}}{n\beta+1}\right]^{\frac{1}{\beta}} \\ &= \frac{1}{n!\,(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha |z-a|^{n+\frac{1}{\beta}},\quad \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1\end{aligned}\tag{1.6}$$

and, obviously,

$$\begin{aligned}M\left(f^{(n+1)};a,z\right) &\le \frac{1}{n!}|z-a|^n\left|\int_a^z \left|f^{(n+1)}(t)\right| dt\right| \\ &\le \frac{1}{n!}|z-a|^n\left\|f^{(n+1)}\right\|_1\end{aligned}\tag{1.7}$$

for all $z,a\in[r,R]$.

Consequently, by (1.4)-(1.7), we may write (see also [**33**] for a similar inequality),

$$\begin{aligned}&\left|f(z)-f(a)-\sum_{k=1}^{n}\frac{(z-a)^k}{k!}f^{(k)}(a)\right| \\ &\le \begin{cases}\frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty |z-a|^{n+1} & \text{if } f^{(n+1)}\in L_\infty[r,R]; \\[2mm] \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha |z-a|^{n+\frac{1}{\beta}} & \text{if } f^{(n+1)}\in L_\alpha[r,R], \\ & \quad \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\[2mm] \frac{1}{n!}\left\|f^{(n+1)}\right\|_1 |z-a|^n, & \end{cases}\end{aligned}\tag{1.8}$$

for all $z,a\in[r,R]$.

If in (1.8) we choose $z = \frac{p_i}{q_i}$, $a = 1$, then we obtain,

$$\left| f\left(\frac{p_i}{q_i}\right) - f(1) - \sum_{k=1}^{n} \frac{\left(\frac{p_i}{q_i} - 1\right)^k}{k!} f^{(k)}(1) \right| \tag{1.9}$$

$$\leq \begin{cases} \frac{1}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty \left| \frac{p_i}{q_i} - 1 \right|^{n+1} \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha \left| \frac{p_i}{q_i} - 1 \right|^{n+\frac{1}{\beta}} \\ \frac{1}{n!} \left\| f^{(n+1)} \right\|_1 \left| \frac{p_i}{q_i} - 1 \right|^n, \end{cases}$$

$$\leq \begin{cases} \frac{1}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty (R-r)^{n+1} \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha (R-r)^{n+\frac{1}{\beta}} \\ \frac{1}{n!} \left\| f^{(n+1)} \right\|_1 (R-r)^n, \end{cases}$$

for $i \in \{1, ..., m\}$.

If we multiply (1.9) by $q_i \geq 0$ and sum over i, and then use the generalised triangle inequality,

$$\left| I_f(p,q) - f(1) - \sum_{k=1}^{n} \frac{f^{(k)}(1)}{k!} \cdot \sum_{i=1}^{m} \frac{(p_i - q_i)^k}{q_i^{k-1}} \right|$$

$$\leq \begin{cases} \frac{1}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty \sum_{i=1}^m \frac{|p_i - q_i|^{n+1}}{q_i^n} \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha \sum_{i=1}^m \frac{|p_i - q_i|^{n+\frac{1}{\beta}}}{q_i^{n+\frac{1}{\beta}-1}} \\ \frac{1}{n!} \left\| f^{(n+1)} \right\|_1 \sum_{i=1}^m \frac{|p_i - q_i|^n}{q_i^{n-1}}, \end{cases}$$

$$\leq \begin{cases} \frac{1}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty (R-r)^{n+1} \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha (R-r)^{n+\frac{1}{\beta}} \\ \frac{1}{n!} \left\| f^{(n+1)} \right\|_1 (R-r)^n, \end{cases}$$

and the inequality (1.1) is proved. ∎

The following theorem also holds [**4**].

THEOREM 1.2 (Barnett et al., 2002 [**4**]). *Let f be as in Theorem 1.1. If $p^{(j)}$, $q^{(j)}$ $(j = 1, 2)$ are probability distributions such that,*

$$r \leq \frac{p_i^{(j)}}{q_i^{(j)}} \leq R, \; i \in \{1, ..., m\} \text{ and } j = 1, 2, \tag{1.10}$$

then

$$r \leq \frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q^{(1)} + (1-\lambda) q^{(2)}} \leq R \text{ for } i \in \{1, ..., m\} \text{ and } \lambda \in [0,1] \tag{1.11}$$

and we have the inequality, for $f^{(n+1)} \in L_\alpha[r,R]$,

$$\begin{aligned}(1.12)\quad &\Bigg| I_f\left(\lambda p^{(1)} + (1-\lambda)p^{(2)}, \lambda q^{(1)} + (1-\lambda)q^{(2)}\right) \\ &\quad - \lambda I_f\left(p^{(1)}, q^{(1)}\right) - (1-\lambda) I_f\left(p^{(2)}, q^{(2)}\right) \\ &\quad - \lambda \sum_{k=1}^{n} \frac{1}{k!} \sum_{i=1}^{m} \frac{(1-\lambda)^k}{\left[q_i^{(1)}\right]^{k-1}} (-1)^k \eta_i^k f^{(k)}\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) \\ &\quad - (1-\lambda) \sum_{k=1}^{n} \frac{1}{k!} \sum_{i=1}^{m} \frac{\lambda^k}{\left[q_i^{(2)}\right]^{k-1}} \eta_i^k f^{(k)}\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right) \Bigg|\end{aligned}$$

$$\leq \begin{cases} \frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty \lambda(1-\lambda) \\ \qquad \times \sum_{i=1}^m |\eta_i|^{n+1}\left[\frac{(1-\lambda)^n}{\left[q_i^{(1)}\right]^n} + \frac{\lambda^n}{\left[q_i^{(2)}\right]^n}\right]; \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha \lambda(1-\lambda) \\ \qquad \times \sum_{i=1}^m |\eta_i|^{n+\frac{1}{\beta}}\left[\frac{(1-\lambda)^{n+\frac{1}{\beta}-1}}{\left[q_i^{(1)}\right]^{n+\frac{1}{\beta}-1}} + \frac{\lambda^{n+\frac{1}{\beta}-1}}{\left[q_i^{(2)}\right]^{n+\frac{1}{\beta}-1}}\right]; \\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_1 \lambda(1-\lambda) \\ \qquad \times \sum_{i=1}^m |\eta_i|^{n}\left[\frac{(1-\lambda)^{n-1}}{\left[q_i^{(1)}\right]^{n-1}} + \frac{\lambda^{n-1}}{\left[q_i^{(2)}\right]^{n-1}}\right], \end{cases}$$

where

$$\eta_i = \eta\left(\lambda, p^{(1)}, p^{(2)}, q^{(1)}, q^{(2)}\right)(i) = \frac{\det\begin{bmatrix} p_i^{(1)} & p_i^{(2)} \\ q_i^{(1)} & q_i^{(2)} \end{bmatrix}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}},$$

for all $\lambda \in [0,1]$ *and* $i \in \{1, ..., m\}$.

PROOF. We use the inequality (1.8) to write,

$$\begin{aligned}(1.13)\quad &\Bigg| f\left(\frac{\lambda p_i^{(1)} + (1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)q_i^{(2)}}\right) - f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) \\ &\quad - \sum_{k=1}^{n} \frac{1}{k!}\left(\frac{\lambda p_i^{(1)} + (1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right)^k f^{(k)}\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)\Bigg|\end{aligned}$$

$$\leq \begin{cases} \frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty \left|\frac{\lambda p_i^{(1)} + (1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right|^{n+1} \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha \left|\frac{\lambda p_i^{(1)} + (1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right|^{n+\frac{1}{\beta}} \\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_1 \left|\frac{\lambda p_i^{(1)} + (1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda)q_i^{(2)}} - \frac{p_i^{(1)}}{q_i^{(1)}}\right|^{n}, \end{cases}$$

and

$$\left| f\left(\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}\right) - f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right) - \sum_{k=1}^{n}\frac{1}{k!}\left(\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right)^k f^{(k)}\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right| \le \begin{cases} \frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty \left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right|^{n+1} \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha \left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right|^{n+\frac{1}{\beta}} \\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_1 \left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right|^{n}. \end{cases} \tag{1.14}$$

If we multiply (1.13) by $\lambda q_i^{(1)}$ and (1.14) by $(1-\lambda)\,q_i^{(2)}$, add them and apply the triangle inequality, we obtain,

$$\begin{aligned} &\left| \left(\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}\right) f\left(\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}\right) - \lambda q_i^{(1)} f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) - (1-\lambda)\,q_i^{(2)} f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right) \right. \\ &\quad - \lambda q_i^{(1)}\sum_{k=1}^{n}\frac{1}{k!}\left(\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}-\frac{p_i^{(1)}}{q_i^{(1)}}\right)^k f^{(k)}\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) \\ &\quad \left. - (1-\lambda)\sum_{k=1}^{n}\frac{1}{k!}\left(\frac{\lambda p_i^{(1)}+(1-\lambda)\,p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)\,q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right)^k f^{(k)}\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)\right| \end{aligned} \tag{1.15}$$

$$\le \begin{cases} \frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty \left[\lambda q_i^{(1)}\left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(1)}}{q_i^{(1)}}\right|^{n+1} + (1-\lambda)\,q_i^{(2)}\left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right|^{n+1}\right] \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha \left[\lambda q_i^{(1)}\left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(1)}}{q_i^{(1)}}\right|^{n+\frac{1}{\beta}} + (1-\lambda)\,q_i^{(2)}\left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right|^{n+\frac{1}{\beta}}\right] \\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_1 \left[\lambda q_i^{(1)}\left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(1)}}{q_i^{(1)}}\right|^{n} + (1-\lambda)\,q_i^{(2)}\left|\frac{\lambda p_i^{(1)}+(1-\lambda)p_i^{(2)}}{\lambda q_i^{(1)}+(1-\lambda)q_i^{(2)}}-\frac{p_i^{(2)}}{q_i^{(2)}}\right|^{n}\right], \end{cases}$$

which is equivalent to

$$\text{(1.16)}\quad \Bigg| \left(\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right) f\left(\frac{\lambda p_i^{(1)} + (1-\lambda) p_i^{(2)}}{\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}}\right) - \lambda q_i^{(1)} f\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right) - (1-\lambda) q_i^{(2)} f\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right)$$
$$- \lambda \sum_{k=1}^{n} \frac{1}{k!} \frac{(1-\lambda)^k (Det_{2,1}(i))^k}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^k \left[q_i^{(1)}\right]^{k-1}} f^{(k)}\left(\frac{p_i^{(1)}}{q_i^{(1)}}\right)$$
$$- (1-\lambda) \sum_{k=1}^{n} \frac{1}{k!} \frac{\lambda^k (Det_{1,2}(i))^k}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^k \left[q_i^{(2)}\right]^{k-1}} f^{(k)}\left(\frac{p_i^{(2)}}{q_i^{(2)}}\right) \Bigg|$$

$$\leq \begin{cases} \frac{1}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty \left[\frac{\lambda(1-\lambda)^{n+1} |Det_{2,1}(i)|^{n+1}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n+1} \left[q_i^{(1)}\right]^n} + \frac{(1-\lambda)\lambda^{n+1} |Det_{2,1}(i)|^{n+1}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n+1} \left[q_i^{(2)}\right]^n} \right]; \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha \left[\frac{\lambda(1-\lambda)^{n+\frac{1}{\beta}} |Det_{2,1}(i)|^{n+\frac{1}{\beta}}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n+\frac{1}{\beta}} \left[q_i^{(1)}\right]^{n+\frac{1}{\beta}-1}} + \frac{(1-\lambda)\lambda^{n+\frac{1}{\beta}} |Det_{2,1}(i)|^{n+\frac{1}{\beta}}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n+\frac{1}{\beta}} \left[q_i^{(2)}\right]^{n+\frac{1}{\beta}-1}} \right]; \\ \frac{1}{n!} \left\| f^{(n+1)} \right\|_1 \left[\frac{\lambda(1-\lambda)^{n} |Det_{2,1}(i)|^{n}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n} \left[q_i^{(1)}\right]^{n-1}} + \frac{(1-\lambda)\lambda^{n} |Det_{2,1}(i)|^{n}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n} \left[q_i^{(2)}\right]^{n-1}} \right]; \end{cases}$$

$$= \begin{cases} \frac{1}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty \frac{\lambda(1-\lambda) |Det_{2,1}(i)|^{n+1}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n+1}} \left[\frac{(1-\lambda)^n}{\left[q_i^{(1)}\right]^n} + \frac{\lambda^n}{\left[q_i^{(2)}\right]^n} \right]; \\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha \frac{\lambda(1-\lambda) |Det_{2,1}(i)|^{n+\frac{1}{\beta}}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n+\frac{1}{\beta}}} \times \left[\frac{(1-\lambda)^{n+\frac{1}{\beta}-1}}{\left[q_i^{(1)}\right]^{n+\frac{1}{\beta}-1}} + \frac{\lambda^{n+\frac{1}{\beta}-1}}{\left[q_i^{(2)}\right]^{n+\frac{1}{\beta}-1}} \right]; \\ \frac{1}{n!} \left\| f^{(n+1)} \right\|_1 \frac{\lambda(1-\lambda) |Det_{2,1}(i)|^{n}}{\left[\lambda q_i^{(1)} + (1-\lambda) q_i^{(2)}\right]^{n}} \left[\frac{(1-\lambda)^{n-1}}{\left[q_i^{(1)}\right]^{n-1}} + \frac{\lambda^{n-1}}{\left[q_i^{(2)}\right]^{n-1}} \right], \end{cases}$$

where

$$Det_{y,z}(i) = \det \begin{bmatrix} p_i^{(y)} & p_i^{(z)} \\ q_i^{(y)} & q_i^{(z)} \end{bmatrix}, \qquad (y,z) \in \{(1,2), (2,1)\}.$$

Summing (1.16) over i and using the generalised triangle inequality, we deduce the desired inequality (1.12). ∎

In [**35**], S.S. Dragomir proved the following perturbed Taylor's formula which is an improvement of a result due to Matić, Pečarić and Ujević [**106**]. It is instructive to give the details here for the sake of completeness.

LEMMA 1.3 (Dragomir, 2000 [**35**]). *Let $f : I \subseteq \mathbb{R} \to \mathbb{R}$ be such that $f^{(n)}$ is absolutely continuous and $f^{(n+1)} \in L_2(I)$, then we have,*

$$f(z) = \sum_{k=0}^{n} \frac{(z-a)^k}{k!} f^{(k)}(a) + \frac{(z-a)^{n+1}}{(n+1)!} \left[f^{(n)}; a, z\right] + G_n(f; a, z) \tag{1.17}$$

for all $z \in I$, where,

$$\left[f^{(n)}; a, z\right] = \frac{f^{(n)}(z) - f^{(n)}(a)}{z-a}$$

and $G_n(f; a, z)$ satisfies the estimate,

$$|G_n(f; a, z)| \le \frac{n(z-a)^{n+1}}{(n+1)!\sqrt{2n+1}} \left[\frac{1}{z-a} \left\|f^{(n+1)}\right\|_2^2 - \left(\left[f^{(n)}; a, z\right]\right)^2\right]^{\frac{1}{2}} \tag{1.18}$$

for all $z \ge a$.

PROOF. Recall Korkine's identity for the mappings h, g,

$$\frac{1}{z-a} \int_a^z h(t) g(t)\, dt - \frac{1}{(z-a)^2} \int_a^z h(t)\, dt \cdot \int_a^z g(t)\, dt = \frac{1}{2(z-a)} \int_a^z \int_a^z (h(t) - h(s))(g(t) - g(s))\, dtds. \tag{1.19}$$

Using (1.19), we have,

$$\int_a^z \frac{(z-t)^n}{n!} f^{(n+1)}(t)\, dt - \frac{1}{z-a} \int_a^z \frac{(z-t)^n}{n!} dt \cdot \int_a^z f^{(n+1)}(t)\, dt = \frac{1}{2(z-a)} \int_a^z \int_a^z \left(\frac{(z-t)^n - (z-s)^n}{n!}\right) \left(f^{(n+1)}(t) - f^{(n+1)}(s)\right) dtds$$

and using Taylor's representation (1.3) and the formula (1.17), we conclude that,

$$G_n(f; a, z) = \frac{1}{2(z-a)} \int_a^z \int_a^z \left[\frac{(z-t)^n - (z-s)^n}{n!}\right] \left(f^{(n+1)}(t) - f^{(n+1)}(s)\right) dtds. \tag{1.20}$$

Now, using the Cauchy-Buniakowski-Schwartz integral inequality for double integrals, we have,

$$|G_n(f; a, z)| \le \frac{1}{2(z-a)} \left[\int_a^z \int_a^z \left[\frac{(z-t)^n - (z-s)^n}{n!}\right]^2 dtds \times \int_a^z \int_a^z \left[f^{(n+1)}(t) - f^{(n+1)}(s)\right]^2 dtds\right]^{\frac{1}{2}}. \tag{1.21}$$

Elementary calculations show that,

$$\frac{1}{2(z-a)^2} \int_a^z \int_a^z \left[\frac{(z-t)^n - (z-s)^n}{n!}\right]^2 dtds = \frac{n^2 (z-a)^{2n}}{[(n+1)!]^2 (2n+1)}$$

and (see also [**106**])

$$\frac{1}{2(z-a)^2}\int_a^z\int_a^z\left[f^{(n+1)}(t)-f^{(n+1)}(s)\right]^2 dtds$$
$$=\frac{1}{z-a}\left\|f^{(n+1)}\right\|_2^2-\left(\left[f^{(n)};a,z\right]\right)^2,$$

and so, by (1.21), we deduce (1.18). ∎

Now, by the Grüss inequality, we may state that,

$$\begin{aligned}0&\le\frac{1}{z-a}\int_a^z\left[f^{(n+1)}(t)\right]^2dt-\left(\frac{1}{z-a}\int_a^z f^{(n+1)}(t)\,dt\right)^2\\&\le\frac{1}{4}\left(\Gamma(z)-\gamma(z)\right),\end{aligned}$$

where

$$\gamma(z)\le f^{(n+1)}(t)\le\Gamma(z)\quad\text{for all } t\in[a,z]. \tag{1.22}$$

By Lemma 1.3, we can obtain the result in [**35**], showing that (1.18) is an improvement on the pre-Grüss result obtained in [**106**].

COROLLARY 1.4. *Let $f:I\subseteq\mathbb{R}\to\mathbb{R}$ be such that $f^{(n)}$ is absolutely continuous and $f^{(n+1)}$ is bounded and satisfies (1.22), then we have the representation (1.17) and the remainder $G_n(f;a,z)$ satisfies the estimate,*

$$|G_n(f;a,z)|\le\frac{n(z-a)^{n+1}}{2(n+1)!\sqrt{2n+1}}\left(\Gamma(z)-\gamma(z)\right) \tag{1.23}$$

for all $z\ge a$.

If $z\le a$, then a similar bound can be stated and so, in general, for any $a\in I$, we have the representation (1.17) and the bounds,

$$\begin{aligned}&|G_n(f;a,z)|\\&\le\frac{n|z-a|^{n+1}}{(n+1)!\sqrt{2n+1}}\left[\frac{\int_a^z\left[f^{(n+1)}(t)\right]^2dt}{z-a}-\left(\left[f^{(n)};a,z\right]\right)^2\right]^{\frac{1}{2}}\\&\le\frac{n|z-a|^{n+1}}{2(n+1)!\sqrt{2n+1}}\left(\Gamma(z)-\gamma(z)\right),\end{aligned} \tag{1.24}$$

where

$$\Gamma:=\sup_{z\in I}f^{(n+1)}(z)<\infty\ \text{ and }\ \gamma:=\inf_{z\in I}f^{(n+1)}(z)>-\infty.$$

In what follows, we use the estimate (1.24) [**4**].

THEOREM 1.5 (Barnett et al., 2002 [**4**]). *Let $f:(0,\infty)\to\mathbb{R}$ be n-time differentiable and such that $f^{(n)}$ is absolutely continuous on $[r,R]$, where $0<r\le1\le R<\infty$. Assume that the probability distributions p,q satisfy the condition,*

$$r\le\frac{p_i}{q_i}\le R\ \text{ a.e on }\Gamma, \tag{1.25}$$

then we have the inequalities,

$$\begin{aligned}(1.26)\qquad &\Bigg| I_f(p,q) - f(1) - \left[\sum_{k=1}^{n} \frac{f^{(k)}(1)}{k!} D_{\chi^k}(p,q)\right] \\ &\quad - \frac{1}{(n+1)!} I_{(\cdot-1)^k f^{(k)}(\cdot)}(p,q) + \frac{f^{(n)}(1)}{(n+1)!} D_{\chi^n}(p,q)\Bigg| \\ &\leq \frac{n}{(n+1)!\sqrt{2n+1}} B\left(p,q,f^{(n+1)}\right) \\ &\leq \frac{n(\Phi-\phi)}{(n+1)!\sqrt{2n+1}} D_{|\chi|^{n+1}}(p,q) \leq \frac{n(\Phi-\phi)}{(n+1)!\sqrt{2n+1}} (R-r)^{n+1},\end{aligned}$$

where

$$\Phi := \sup_{z\in[r,R]} f^{(n+1)}(z) < \infty \ \textit{and}\ \phi := \inf_{z\in[r,R]} f^{(n+1)}(z) > -\infty$$

and

$$B\left(p,q,f^{(n+1)}\right) := I_g(p,q)$$

where

$$g(z) = |z-1|^{n+1}\left[\frac{1}{z-1}\int_1^z \left[f^{(n+1)}(t)\right]^2 dt - \left(\frac{f^{(n)}(z)-f^{(n)}(1)}{z-1}\right)^2\right]^{\frac{1}{2}}.$$

PROOF. Apply the inequality (1.24) for $a = 1$ and $z = \frac{p_i}{q_i}$ to obtain,

$$\begin{aligned}&\Bigg| f\left(\frac{p_i}{q_i}\right) - f(1) - \sum_{k=1}^{n} \frac{f^{(k)}(1)}{k!}\left(\frac{p_i}{q_i}-1\right)^k \\ &\qquad - \frac{\left(\frac{p_i}{q_i}-1\right)^n}{(n+1)!}\left[f^{(n)}\left(\frac{p_i}{q_i}\right) - f^{(n)}(1)\right]\Bigg| \\ &\leq \frac{n\left|\frac{p_i}{q_i}-1\right|^n}{(n+1)!\sqrt{2n+1}} \frac{q_i}{p_i-q_i}\int_1^{\frac{p_i}{q_i}} \left[f^{(n+1)}(t)\right]^2 dt \\ &\qquad - \left(\frac{f^{(n)}\left(\frac{p_i}{q_i}\right) - f^{(n)}(1)}{\frac{p_i}{q_i}-1}\right)^2\Bigg]^{\frac{1}{2}} \\ &\leq \frac{n\left|\frac{p_i}{q_i}-1\right|^{n+1}(\Phi-\phi)}{(n+1)!\sqrt{2n+1}}.\end{aligned}$$

If we multiply by $p_i \geq 0$, sum over i and use the generalised triangle inequality, we deduce,

$$\begin{aligned}&\Bigg| I_f(p,q) - f(1) - \sum_{k=1}^{n} \frac{f^{(k)}(1)}{k!} D_{\chi^k}(p,q) \\ &\qquad - \frac{1}{(n+1)!}\sum_{i=1}^{m} q_i\left(\frac{p_i}{q_i}-1\right)^k f^{(k)}\left(\frac{p_i}{q_i}\right) + \frac{f^{(n)}(1)}{(n+1)!} D_{\chi^n}(p,q)\Bigg|\end{aligned}$$

$$\leq \frac{n(\Phi-\phi)}{(n+1)!\sqrt{2n+1}} \sum_{i=1}^{m} q_i \left|\frac{p_i}{q_i}-1\right|^{n+1} \left[\frac{q_i}{p_i-q_i}\int_1^{\frac{p_i}{q_i}} \left[f^{(n+1)}(t)\right]^2 dt \right.$$
$$\left. -q_i^2 \frac{\left|f^{(n)}\left(\frac{p_i}{q_i}\right)-f^{(n)}(1)\right|^2}{(p_i-q_i)^2}\right]^{\frac{1}{2}}$$
$$\leq \frac{n(\Phi-\phi)}{2(n+1)!\sqrt{2n+1}} D_{|\chi|^{n+1}}(p,q) \leq \frac{n(\Phi-\phi)}{2(n+1)!\sqrt{2n+1}}(R-r)^{n+1}$$

and the theorem is proved. ∎

2. Some Particular Inequalities

The following proposition holds [**4**].

PROPOSITION 2.1 (Barnett et al., 2002 [**4**]). *Let p,q be two probability distributions satisfying the condition*

$$0 < r \leq \frac{p_i}{q_i} \leq R < \infty \text{ a.e on } \Gamma, \tag{2.1}$$

then, for $n \geq 1$, we have the inequality,

$$\left|D_{KL}(q,p) - \sum_{k=1}^{n} \frac{(-1)^k}{k!} D_{\chi^k}(p,q)\right| \tag{2.2}$$

$$\leq \begin{cases} \frac{1}{(n+1)r^{n+1}} D_{|\chi|^{n+1}}(p,q); \\ \frac{1}{(n\beta+1)^{\frac{1}{\beta}}[(n+1)\alpha-1]^{\frac{1}{\alpha}}} \left[\frac{R^{(n+1)\alpha-1}-r^{(n+1)\alpha-1}}{R^{(n+1)\alpha-1}r^{(n+1)\alpha-1}}\right]^{\frac{1}{\alpha}} D_{|\chi|^{n+\frac{1}{\beta}}}(p,q), \\ \qquad \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{1}{n}\cdot\frac{R^n-r^n}{R^n r^n} D_{|\chi|^n}(p,q); \end{cases}$$

$$\leq \begin{cases} \frac{1}{(n+1)r^{n+1}} (R-r)^{n+1}; \\ \frac{1}{(n\beta+1)^{\frac{1}{\beta}}[(n+1)\alpha-1]^{\frac{1}{\alpha}}} \left[\frac{R^{(n+1)\alpha-1}-r^{(n+1)\alpha-1}}{R^{(n+1)\alpha-1}r^{(n+1)\alpha-1}}\right]^{\frac{1}{\alpha}} (R-r)^{n+\frac{1}{\beta}}, \\ \qquad \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{1}{n}\cdot\frac{R^n-r^n}{R^n r^n} (R-r)^n; \end{cases}$$

PROOF. Consider the mapping $f(t)=\ln t$. We have,

$$I_f(p,q) = \sum_{i=1}^{m} q_i f\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{m} q_i \ln\left(\frac{p_i}{q_i}\right)$$
$$= -\sum_{i=1}^{m} q_i \ln\left(\frac{q_i}{p_i}\right) = -D_{KL}(q,p),$$

$$f^{(k)}(t) = \frac{(-1)^{k-1}(k-1)!}{t^k}, \quad k\in\mathbb{N}, k\geq 1$$

for $\alpha > 1$ and

$$\left\|f^{(n+1)}\right\|_{\infty} := \sup_{t\in[r,R]} \left|f^{(n+1)}(t)\right| = n! \sup_{t\in[r,R]} \left\{\frac{1}{t^{n+1}}\right\} = \frac{n!}{r^{n+1}};$$

$$\begin{aligned}\left\|f^{(n+1)}\right\|_{\alpha} &:= \left(\int_r^R \left|f^{(n+1)}(t)\right|^{\alpha} dt\right)^{\frac{1}{\alpha}} = n! \left[\int_r^R \frac{dt}{t^{(n+1)\alpha}}\right]^{\frac{1}{\alpha}} \\ &= n! \left[\frac{t^{-(n+1)\alpha+1}}{-(n+1)\alpha+1} dt\bigg|_r^R\right]^{\frac{1}{\alpha}} \\ &= n! \left[\frac{R^{(n+1)\alpha-1} - r^{(n+1)\alpha-1}}{[(n+1)\alpha-1] R^{(n+1)\alpha-1} r^{(n+1)\alpha-1}}\right]^{\frac{1}{\alpha}}.\end{aligned}$$

Applying Theorem 1.1 and using the above assumptions, we deduce the desired inequality (2.2). ∎

The following proposition also holds [**4**].

PROPOSITION 2.2 (Barnett et al., 2002 [**4**]). *Let* p, q *be as in the above Proposition 2.1, then we have the inequality,*

$$(2.3)\qquad \left|D_{KL}(q,p) - \sum_{k=2}^{n} \frac{(-1)^k}{(k-1)k} D_{\chi^k}(p,q)\right|$$
$$\leq \begin{cases} \frac{1}{n(n+1)r^{n+1}} D_{|\chi|^{n+1}}(p,q); \\ \frac{1}{n(n\beta+1)^{\frac{1}{\beta}}(n\alpha-1)^{\frac{1}{\alpha}}} \left[\frac{R^{n\alpha-1}-r^{n\alpha-1}}{R^{n\alpha-1}r^{n\alpha-1}}\right]^{\frac{1}{\alpha}} D_{|\chi|^{n+\frac{1}{\beta}}}(p,q), \\ \qquad\qquad \alpha > 1,\ \frac{1}{\alpha}+\frac{1}{\beta} = 1; \\ \frac{1}{(n-1)n} \cdot \frac{R^{n-1}-r^{n-1}}{R^{n-1}r^{n-1}} D_{|\chi|^n}(p,q); \end{cases}$$
$$\leq \begin{cases} \frac{1}{n(n+1)r^{n+1}} (R-r)^{n+1}; \\ \frac{1}{n(n\beta+1)^{\frac{1}{\beta}}(n\alpha-1)^{\frac{1}{\alpha}}} \left[\frac{R^{n\alpha-1}-r^{n\alpha-1}}{R^{n\alpha-1}r^{n\alpha-1}}\right]^{\frac{1}{\alpha}} (R-r)^{n+\frac{1}{\beta}}, \\ \qquad\qquad \alpha > 1,\ \frac{1}{\alpha}+\frac{1}{\beta} = 1; \\ \frac{1}{(n-1)n} \cdot \frac{R^{n-1}-r^{n-1}}{R^{n-1}r^{n-1}} (R-r)^n. \end{cases}$$

PROOF. Consider the mapping $f(t) = t\ln(t)$. We have,

$$\begin{aligned}I_f(p,q) &= \sum_{i=1}^{m} q_i f\left(\frac{p_i}{q_i}\right) = \sum_{i=1}^{m} q_i \frac{p_i}{q_i} \ln\left(\frac{p_i}{q_i}\right) \\ &= \sum_{i=1}^{m} p_i \ln\left(\frac{p_i}{q_i}\right) = D_{KL}(p,q),\end{aligned}$$

$$\begin{aligned} f^{(1)}(t) &= \ln t + 1, \\ f^{(k)}(t) &= \frac{(-1)^k (k-2)!}{t^{k-1}}, \quad k \geq 2 \\ \left\| f^{(n+1)} \right\|_\infty &= \frac{(n-1)!}{r^n}, \\ \left\| f^{(n+1)} \right\|_\alpha &= (n-1)! \left[\frac{R^{n\alpha-1} - r^{n\alpha-1}}{R^{n\alpha-1} r^{n\alpha-1}} \right]^{\frac{1}{\alpha}}, \ \alpha > 1 \end{aligned}$$

and

$$\left\| f^{(n+1)} \right\|_1 = (n-2)! \frac{R^{n-1} - r^{n-1}}{R^{n-1} r^{n-1}}.$$

Applying Theorem 1.1 for the mapping $f(t) = t \ln t$, we have,

$$\begin{aligned} &\left| D_{KL}(p,q) - f^{(1)}(1) D_\chi(p,q) - \sum_{k=2}^{n} \frac{(-1)^k (k-2)!}{k!} D_{\chi^k}(p,q) \right| \\ &\qquad \leq \begin{cases} \frac{1}{(n+1)!} \frac{(n-1)!}{r^n} D_{|\chi|^{n+1}}(p,q); \\ \frac{1}{n(n\beta+1)^{\frac{1}{\beta}}} \cdot \frac{(n-1)!}{(n\alpha-1)^{\frac{1}{\alpha}}} \left[\frac{R^{n\alpha-1} - r^{n\alpha-1}}{R^{n\alpha-1} r^{n\alpha-1}} \right]^{\frac{1}{\alpha}} D_{|\chi|^{n+\frac{1}{\beta}}}(p,q); \\ \frac{1}{n!} (n-2)! \frac{R^{n-1} - r^{n-1}}{R^{n-1} r^{n-1}} D_{|\chi|^n}(p,q), \end{cases} \end{aligned}$$

i.e.,

$$\begin{aligned} &\left| D_{KL}(q,p) - \sum_{k=2}^{n} \frac{(-1)^k}{(k-1)k} D_{\chi^k}(p,q) \right| \\ &\qquad \leq \begin{cases} \frac{1}{n(n+1)r^{n+1}} D_{|\chi|^{n+1}}(p,q); \\ \frac{1}{n(n\beta+1)^{\frac{1}{\beta}} (n\alpha-1)^{\frac{1}{\alpha}}} \left[\frac{R^{n\alpha-1} - r^{n\alpha-1}}{R^{n\alpha-1} r^{n\alpha-1}} \right]^{\frac{1}{\alpha}} D_{|\chi|^{n+\frac{1}{\beta}}}(p,q); \\ \frac{1}{(n-1)n} \frac{R^{n-1} - r^{n-1}}{R^{n-1} r^{n-1}} D_{|\chi|^n}(p,q) \end{cases} \end{aligned}$$

and the first inequality in (2.3) is proved.

The second inequality is obvious and we omit the details. ∎

REMARK 2.1. Similar results can be obtained if we apply Theorem 1.1 for other particular mappings f, generating the Hellinger, Jeffrey's, Bhattacharyya, or other divergence measures as considered in the introduction.

3. APPROXIMATING f-DIVERGENCES VIA A GENERALISED TAYLOR FORMULA

We may state the following representation result which is a reformulation of Theorem 1 in [**106**]:

THEOREM 3.1. *Let $\{S_n(\cdot,\cdot)\}_{n\in\mathbb{N}}$ be a sequence of polynomials of two variables satisfying the condition:*

$$\frac{\partial S_n(t,x)}{\partial t} = S_{n-1}(t,x), \quad S_0(t,x) = 1 \text{ for } x, t \in \mathbb{R} \text{ and } n \in \mathbb{N}, \tag{3.1}$$

then we have the identity,

$$f(x) = f(a) + \sum_{k=1}^{n} (-1)^{k+1} \left[S_k(x,x) f^{(k)}(x) - S_k(a,x) f^{(k)}(a)\right] + R_n(f;a,x), \tag{3.2}$$

where

$$R_n(f;a,x) := (-1)^n \int_a^x S_n(t,x) f^{(n+1)}(t)\, dt$$

and $f : I \to \mathbb{R}$ is such that $f^{(n)}$ is absolutely continuous on I.

(1) If in (3.2) we set $S_n(t,x) = \frac{1}{n!}(t-x)^n \quad (n \in \mathbb{N})$, we get the Taylor's identity [**33**]:

$$f(x) = f(a) + \sum_{k=1}^{n} \frac{(x-a)^k}{k!} f^{(k)}(a) + \frac{1}{n!}\int_a^x (x-t)^n f^{(n+1)}(t)\, dt, \quad x \in I. \tag{3.3}$$

(2) If in (3.2) we set $S_n(t,x) = \frac{1}{n!}\left(t - \frac{a+x}{2}\right)^n, \quad (n \in \mathbb{N})$, then we have the identity [**106**]:

$$f(x) = f(a) + \sum_{k=1}^{n} \frac{(x-a)^k}{2^k k!} \left[f^{(k)}(a) + (-1)^{k+1} f^{(k)}(x)\right] + \frac{(-1)^n}{n!}\int_a^x \left(x - \frac{a+x}{2}\right)^n f^{(n+1)}(t)\, dt, \quad x \in I. \tag{3.4}$$

(3) If in (3.2) we set $S_n(t,x) = \frac{(x-a)^n}{n!} B_n\left(\frac{t-a}{x-a}\right)$, $n \in \mathbb{N}$, $S_0(t,x) = 1$, where B_n denotes the Bernoulli polynomial and $B_n := B_n(0)$ are the Bernoulli numbers, then we have the following representation [**106**],

$$\begin{aligned} f(x) = f(a) &+ \frac{x-a}{2}\left[f'(x) + f'(a)\right] \\ &- \sum_{k=1}^{\left[\frac{n}{2}\right]} \frac{(x-a)^{2k}}{(2k)!} B_{2k}\left[f^{(2k)}(x) - f^{(2k)}(a)\right] \\ &+ (-1)^n \frac{(x-a)^n}{n!}\int_a^x B_n\left(\frac{t-a}{x-a}\right) f^{(n+1)}(t)\, dt. \end{aligned} \tag{3.5}$$

(4) If in (3.2) we set $S_n(t,x) = \frac{(x-a)^n}{n!} E_n\left(\frac{t-a}{x-a}\right)$, $n \in \mathbb{N}$, $S_0(t,x) = 1$, where $E_n(t)$ denotes the Euler polynomials and B_n are the Bernoulli numbers, then we have the representation [**106**],

$$\begin{aligned} f(x) = f(a) &+ 2\sum_{k=1}^{\left[\frac{n+1}{2}\right]} \frac{(x-a)^{2k-1}\left(4^k - 1\right)}{(2k)!} \\ &\times B_{2k}\left[f^{(2k-1)}(x) + f^{(2k-1)}(a)\right] \\ &+ (-1)^n \frac{(x-a)^n}{n!}\int_a^x E_n\left(\frac{t-a}{x-a}\right) f^{(n+1)}(t)\, dt. \end{aligned} \tag{3.6}$$

We are now able to point out the following representation for the $f-$divergence [**64**].

THEOREM 3.2. *Let* $\{S_n(t,z)\}_{n\in\mathbb{N}}$ *and* $f: I \subseteq \mathbb{R} \to \mathbb{R}$ *be as in Theorem 3.1. If* $p, q \in \Omega$*, then we have the representation,*

$$I_f(p,q) = f(1) + \sum_{k=1}^{n} (-1)^{k+1} \left[I_{S_k(\cdot,\cdot)f^{(k)}(\cdot)}(p,q) - I_{S_k(1,\cdot)f^{(k)}(1)}(p,q) \right] + R_f(p,q), \tag{3.7}$$

where the remainder $R_f(p,q)$ *can be given by,*

$$R_f(p,q) = (-1)^n \sum_{i=1}^{m} q_i \left(\int_1^{\frac{p_i}{q_i}} S_n\left(t, \frac{p_i}{q_i}\right) f^{(n+1)}(t)\,dt \right). \tag{3.8}$$

PROOF. From the representation (3.2), we may write,

$$f\left(\frac{p_i}{q_i}\right) = f(1) + \sum_{k=1}^{n} (-1)^{k+1} \left[S_k\left(\frac{p_i}{q_i}, \frac{p_i}{q_i}\right) f^{(k)}\left(\frac{p_i}{q_i}\right) - S_k\left(1, \frac{p_i}{q_i}\right) f^{(k)}(1) \right] + (-1)^n \int_1^{\frac{p_i}{q_i}} S_n\left(t, \frac{p_i}{q_i}\right) f^{(n+1)}(t)\,dt \tag{3.9}$$

for all $i \in \{1, ..., m\}$.

If we multiply (3.9) by $q_i \geq 0$, sum over i and take into account that $\sum_{i=1}^{m} q_i = 1$, then we obtain the desired representation (3.7). ∎

The following particular cases are important in applications.

(1) If we use the representation (3.3), we get,

$$I_f(p,q) = f(1) + \sum_{k=1}^{n} \frac{f^{(k)}(1)}{k!} D_k(p,q) + \frac{1}{n!} \sum_{i=1}^{m} q_i \left[\int_1^{\frac{p_i}{q_i}} \left(\frac{p_i}{q_i} - t \right)^n f^{(n+1)}(t)\,dt \right] \tag{3.10}$$

for all p, q, where,

$$D_k(p,q) := \sum_{i=1}^{m} q_i^{-k+1} (p_i - q_i)^k, \quad k = 1, \ldots, n.$$

(2) From the identity (3.4),

$$I_f(p,q) = f(1) + \sum_{k=1}^{n} \frac{f^{(k)}(1)}{2^k k!} D_k(p,q) + \sum_{k=1}^{n} \frac{(-1)^{k+1}}{2^k k!} I_{(\cdot-1)^k f^{(k)}(\cdot)}(p,q) + \frac{1}{n!} \sum_{i=1}^{m} q_i \left[\int_1^{\frac{p_i}{q_i}} \left(1 - \frac{p_i + q_i}{q_i} \right)^n f^{(n+1)}(t)\,dt \right] \tag{3.11}$$

for all p, q.

(3) If we use the identity (3.5), we may obtain:

$$I_f(p,q) = f(1) + \sum_{i=1}^{m}\left(\frac{p_i - q_i}{2}\right) f'\left(\frac{p_i}{q_i}\right) + \sum_{k=1}^{\left[\frac{n}{2}\right]} \frac{f^{(2k)}(1)}{(2k)!} D_{2k}(p,q) - \sum_{k=1}^{\left[\frac{n}{2}\right]} \frac{B_{2k}}{(2k)!} I_{(\cdot-1)^{2k} f^{(2k)}(\cdot)}(p,q) + \frac{(-1)^n}{n!}\sum_{i=1}^{m} q_i^{-n+1}(p_i - q_i)^n \times \left[\int_1^{\frac{p_i}{q_i}} B_n\left(\frac{t-1}{\frac{p_i}{q_i}-1}\right) f^{(n+1)}(t)\,dt\right] \tag{3.12}$$

for all p, q.

(4) Finally, by use of identity (3.6), we may write,

$$I_f(p,q) = f(1) + 2\sum_{k=1}^{\left[\frac{n+1}{2}\right]} \frac{\left(4^k - 1\right) B_{2k} f^{(k)}(1)}{(2k)!} D_{2k-1}(p,q) + 2\sum_{k=1}^{\left[\frac{n+1}{2}\right]} \frac{B_{2k}\left(4^k-1\right)}{(2k)!} I_{(\cdot-1)^{2k-1} f^{(2k-1)}}(p,q) + \frac{(-1)^n}{n!}\sum_{i=1}^{m} q_i^{-n+1}(p_i - q_i)^n \times \left[\int_1^{\frac{p_i}{q_i}} E_n\left(\frac{t-1}{\frac{p_i}{q_i}-1}\right) f^{(n+1)}(t)\,dt\right] \tag{3.13}$$

for all p, q.

4. Bounds for the Remainder

For $a, b \in \mathbb{R}$, we denote

$$\|f\|_{[a,b],p} := \left|\int_a^b |f(t)|^p\,dt\right|^{\frac{1}{p}} \quad \text{if } p \in [1,\infty)$$

and

$$\|f\|_{[a,b],\infty} := ess \sup_{\substack{t\in[a,b] \\ (t\in[b,a])}} |f(t)|.$$

It is obvious that the order $a < b$ or $a > b$ is irrelevant in the definitions of the above Lebesgue $p-$norms.

The following general theorem involving the estimation of the remainder $R_f(p,q)$ holds [**64**].

THEOREM 4.1 (Dragomir & Gluščević, 2001 [**64**]). *Assume that* $\{S_n(t,x)\}_{n\in\mathbb{N}}$ *and* f *are as in Theorem 3.1. If* $p,q\in\Omega$*, then we have the inequality,*

$$|R_f(p,q)| \le \begin{cases} \sum_{i=1}^m q_i \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],\infty} \times \left\|S_n\left(\cdot,\frac{p_i}{q_i}\right)\right\|_{\left[1,\frac{p_i}{q_i}\right],1}, \\ \sum_{i=1}^m q_i \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],\alpha} \times \left\|S_n\left(\cdot,\frac{p_i}{q_i}\right)\right\|_{\left[1,\frac{p_i}{q_i}\right],\beta}, \\ \qquad \text{if } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1, \\ \sum_{i=1}^m q_i \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],1} \times \left\|S_n\left(\cdot,\frac{p_i}{q_i}\right)\right\|_{\left[1,\frac{p_i}{q_i}\right],\infty}. \end{cases} \tag{4.1}$$

PROOF. We have that

$$|R_f(p,q)| \le \sum_{i=1}^m q_i \left|\int_1^{\frac{p_i}{q_i}} S_n\left(t,\frac{p_i}{q_i}\right) f^{(n+1)}(t)\,dt\right|. \tag{4.2}$$

Now, observe that,

$$\left|\int_1^{\frac{p_i}{q_i}} S_n\left(t,\frac{p_i}{q_i}\right) f^{(n+1)}(t)\,dt\right| \le \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],\infty} \times \left\|S_n\left(\cdot,\frac{p_i}{q_i}\right)\right\|_{\left[1,\frac{p_i}{q_i}\right],1} \tag{4.3}$$

and, by Hölder's inequality for $\alpha>1$, $\frac{1}{\alpha}+\frac{1}{\beta}=1$,

$$\begin{aligned} &\left|\int_1^{\frac{p_i}{q_i}} S_n\left(t,\frac{p_i}{q_i}\right) f^{(n+1)}(t)\,dt\right| \\ &\le \left|\int_1^{\frac{p_i}{q_i}} \left|S_n\left(t,\frac{p_i}{q_i}\right)\right|^\beta dt\right|^{\frac{1}{\beta}} \times \left|\int_1^{\frac{p_i}{q_i}} \left|f^{(n+1)}(t)\right|^\alpha dt\right|^{\frac{1}{\alpha}} \\ &= \left\|S_n\left(\cdot,\frac{p_i}{q_i}\right)\right\|_{\left[1,\frac{p_i}{q_i}\right],\beta} \times \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],\alpha}. \end{aligned} \tag{4.4}$$

Finally,

$$\left|\int_1^{\frac{p_i}{q_i}} S_n\left(t,\frac{p_i}{q_i}\right) f^{(n+1)}(t)\,dt\right| \le \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],1} \times \left\|S_n\left(\cdot,\frac{p_i}{q_i}\right)\right\|_{\left[1,\frac{p_i}{q_i}\right],\infty} \tag{4.5}$$

for all $i\in\{1,...,m\}$.

Using (4.2) and (4.3) – (4.5), we deduce (4.1). ∎

REMARK 4.1. If we assume that $0 \le r \le \frac{p_i}{q_i} \le R < \infty$ for $i \in \{1, ..., m\}$, then obviously $r \le 1 \le R$ and the right side of the inequality (4.1) may be upper bounded by,

$$\begin{cases} \left\| f^{(n+1)} \right\|_{[r,R],\infty} \sum_{i=1}^{m} q_i \left\| S_n \left(\cdot, \frac{p_i}{q_i} \right) \right\|_{\left[1, \frac{p_i}{q_i}\right],1}, \\ \left\| f^{(n+1)} \right\|_{[r,R],\alpha} \sum_{i=1}^{m} q_i \left\| S_n \left(\cdot, \frac{p_i}{q_i} \right) \right\|_{\left[1, \frac{p_i}{q_i}\right],\beta}, \\ \left\| f^{(n+1)} \right\|_{[r,R],1} \sum_{i=1}^{m} q_i \left\| S_n \left(\cdot, \frac{p_i}{q_i} \right) \right\|_{\left[1, \frac{p_i}{q_i}\right],\infty}, \end{cases}$$

If we choose some particular instances of polynomials $S_n(\cdot, \cdot)$ we may compute the Lebesgue norm $\left\| S_n \left(\cdot, \frac{p_i}{q_i} \right) \right\|_s$, $s \in [1, \infty]$, obtaining more explicit bounds for the remainder $R_f(p, q)$.

(1) If we choose $S_n(t, z) = \frac{1}{n!}(t - z)^n$, then,

$$\|S_n(\cdot, z)\|_{[1,z],1} = \frac{1}{n!} \left| \int_1^z |t - z|^n \, dt \right| = \frac{1}{(n+1)!} |z - 1|^{n+1},$$

$$\begin{aligned} \|S_n(\cdot, z)\|_{[1,z],\alpha} &= \frac{1}{n!} \left| \int_1^z |t - z|^{\alpha n} \, dt \right|^{\frac{1}{\alpha}} \\ &= \frac{1}{n!} \left[\frac{|z - 1|^{\alpha n + 1}}{\alpha n + 1} \right]^{\frac{1}{\alpha}} = \frac{|z - 1|^{n + \frac{1}{\alpha}}}{n! (\alpha n + 1)^{\frac{1}{\alpha}}}, \end{aligned}$$

and

$$\|S_n(\cdot, z)\|_{[1,z],\infty} = \frac{1}{n!} |z - 1|^n.$$

Consequently, we may state the following corollary which is useful in practice.

COROLLARY 4.2. *Let f be as in Theorem 3.1, then, for p, q two probability distributions, we have*

$$I_f(p, q) = f(1) + \sum_{k=1}^{n} \frac{f^{(k)}(1)}{k!} D_k(p, q) + R_f(p, q), \tag{4.6}$$

where

$$D_k(p, q) := \sum_{i=1}^{m} q_i^{-k+1} (p_i - q_i)^k$$

and the remainder $R_f(p, q)$ satisfies the bound

$$|R_f(p, q)| \tag{4.7}$$

$$\le \begin{cases} \frac{1}{(n+1)!} \sum_{i=1}^{m} |p_i - q_i|^{n+1} q_i^{-n} \left\| f^{(n+1)} \right\|_{\left[1, \frac{p_i}{q_i}\right],\infty}, \\ \frac{1}{n!(\alpha n+1)^{\frac{1}{\alpha}}} \sum_{i=1}^{m} |p_i - q_i|^{n+\frac{1}{\alpha}} q_i^{-n-\frac{1}{\alpha}+1} \left\| f^{(n+1)} \right\|_{\left[1, \frac{p_i}{q_i}\right],\beta}, \\ \qquad\qquad \text{if } \alpha > 1, \ \frac{1}{\alpha} + \frac{1}{\beta} = 1, \\ \frac{1}{n!} \sum_{i=1}^{m} |p_i - q_i|^n q_i^{-n+1} \left\| f^{(n+1)} \right\|_{\left[1, \frac{p_i}{q_i}\right],1}. \end{cases}$$

Moreover, if $0 \leq r \leq \frac{p_i}{q_i} \leq R < \infty$ *for* $i \in \{1, ..., m\}$*, then the right hand side of (4.7) can be upper bounded by,*

$$(4.8)\quad \begin{cases} \frac{\left\|f^{(n+1)}\right\|_{[r,R],\infty}}{(n+1)!} \sum_{i=1}^{m} |p_i - q_i|^{n+1} q_i^{-n}, \\ \frac{\left\|f^{(n+1)}\right\|_{[r,R],\beta}}{n!(\alpha n+1)^{\frac{1}{\alpha}}} \sum_{i=1}^{m} |p_i - q_i|^{n+\frac{1}{\alpha}} q_i^{-n-\frac{1}{\alpha}+1}, \\ \qquad \text{if } \alpha > 1, \ \frac{1}{\alpha} + \frac{1}{\beta} = 1, \\ \frac{\left\|f^{(n+1)}\right\|_{[r,R],1}}{n!} \sum_{i=1}^{m} |p_i - q_i|^{n} q_i^{-n+1}, \end{cases}$$

$$\leq \begin{cases} \frac{\left\|f^{(n+1)}\right\|_{[r,R],\infty}(R-r)^{n+1}}{(n+1)!}, \\ \frac{\left\|f^{(n+1)}\right\|_{[r,R],\beta}(R-r)^{n+\frac{1}{\alpha}}}{n!(\alpha n+1)^{\frac{1}{\alpha}}}, \\ \frac{\left\|f^{(n+1)}\right\|_{[r,R],1}(R-r)^{n}}{n!}. \end{cases}$$

(2) If we choose $S_n(t,z) = \frac{1}{n!}\left(t - \frac{1+z}{2}\right)^n$, then,

$$\|S_n(\cdot,z)\|_{[1,z],1} = \frac{1}{n!}\left|\int_1^z \left|t - \frac{1+z}{2}\right|^n dt\right|.$$

If we assume that $z \geq 1$,

$$\begin{aligned} \int_1^z \left|t - \frac{1+z}{2}\right|^n dt &= \int_1^{\frac{1+z}{2}} \left(\frac{1+z}{2} - t\right)^n dt + \int_{\frac{1+z}{2}}^z \left(t - \frac{1+z}{2}\right)^n dt \\ &= \frac{1}{n+1}\left(\frac{z-1}{2}\right)^{n+1} + \frac{1}{n+1}\left(\frac{z-1}{2}\right)^{n+1} \\ &= \frac{(z-1)^{n+1}}{(n+1)\,2^n}. \end{aligned}$$

If we assume that $z \leq 1$, then,

$$\int_z^1 \left|t - \frac{1+z}{2}\right|^n dt = \frac{(1-z)^{n+1}}{(n+1)\,2^n}$$

and thus, we may state that,

$$\|S_n(\cdot,z)\|_{[1,z],1} = \frac{1}{(n+1)!} \cdot \frac{|z-1|^{n+1}}{2^n}.$$

Similarly, we have,

$$\begin{aligned} \|S_n(\cdot,z)\|_{[1,z],\alpha} &= \frac{1}{n!}\left|\int_1^z \left|t - \frac{1+z}{2}\right|^{n\alpha} dt\right|^{\frac{1}{\alpha}} = \frac{1}{n!}\left[\frac{|z-1|^{n\alpha+1}}{(n\alpha+1)\,2^{n\alpha}}\right]^{\frac{1}{\alpha}} \\ &= \frac{1}{n!} \cdot \frac{|z-1|^{n+\frac{1}{\alpha}}}{(n\alpha+1)^{\frac{1}{\alpha}}\,2^n}, \quad \alpha \geq 1 \end{aligned}$$

and

$$\|S_n(\cdot,z)\|_{[1,z],\infty} = \frac{1}{n!} \cdot \frac{|z-1|^n}{2^n}.$$

Consequently, we may state the following corollary.

COROLLARY 4.3. *Let f be as in Theorem 3.1. Then, for for any p, q we have*

$$I_f(p,q) = f(1) + \sum_{k=1}^{n} \frac{f^{(k)}(1)}{2^k k!} D_k(p,q) + \sum_{k=1}^{n} \frac{(-1)^{k+1}}{2^k k!} I_{(\cdot-1)^k f^{(k)}(\cdot)}(p,q) + \tilde{R}_f(p,q) \tag{4.9}$$

and the remainder $\tilde{R}_f(p,q)$ satisfies the bound

$$\left|\tilde{R}_f(p,q)\right| \le \begin{cases} \frac{1}{(n+1)!2^n} \sum_{i=1}^m q_i^{-n} |p_i - q_i|^{n+1} \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],\infty}, \\ \frac{1}{n!2^n(\alpha n+1)^{\frac{1}{\alpha}}} \sum_{i=1}^m q_i^{-n+1-\frac{1}{\alpha}} |p_i - q_i|^{n+\frac{1}{\alpha}} \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],\beta}, \\ \qquad \text{if } \alpha > 1, \ \frac{1}{\alpha}+\frac{1}{\beta}=1, \\ \frac{1}{n!2^n} \sum_{i=1}^m q_i^{-n+1} |p_i - q_i|^{n} \left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],1}. \end{cases} \tag{4.10}$$

Moreover, if $0 \le r \le \frac{p_i}{q_i} \le R < \infty$ for $i \in \{1, ..., m\}$, then the right hand side of (4.10) can be upper bounded by:

$$\begin{cases} \frac{\left\|f^{(n+1)}\right\|_{[r,R],\infty}}{(n+1)!2^n} \sum_{i=1}^m |p_i - q_i|^{n+1} q_i^{-n}, \\ \frac{\left\|f^{(n+1)}\right\|_{[r,R],\beta}}{2^n n!(\alpha n+1)^{\frac{1}{\alpha}}} \sum_{i=1}^m |p_i - q_i|^{n+\frac{1}{\alpha}} q_i^{-n-\frac{1}{\alpha}+1}, \\ \qquad \text{if } \alpha > 1, \ \frac{1}{\alpha}+\frac{1}{\beta}=1, \\ \frac{\left\|f^{(n+1)}\right\|_{[r,R],1}}{2^n n!} \sum_{i=1}^m |p_i - q_i|^{n} q_i^{-n+1}, \end{cases} \le \begin{cases} \frac{\left\|f^{(n+1)}\right\|_{[r,R],\infty}(R-r)^{n+1}}{(n+1)!2^n}, \\ \frac{\left\|f^{(n+1)}\right\|_{[r,R],\beta}(R-r)^{n+\frac{1}{\alpha}}}{n!(\alpha n+1)^{\frac{1}{\alpha}}2^n}, \quad \alpha > 1, \ \frac{1}{\alpha}+\frac{1}{\beta}=1, \\ \frac{\left\|f^{(n+1)}\right\|_{[r,R],1}(R-r)^{n}}{n!}. \end{cases} \tag{4.11}$$

CHAPTER 17

Approximations Via Some Integral Identities

1. Representation of Csiszár f-Divergence

In [**16**] (see also [**15**]), the authors proved the following integral identity generalising the mid-point rule.

LEMMA 1.1. *Let* $g:[a,b]\to\mathbb{R}$ *be a function such that* $g^{(n-1)}$ *is absolutely continuous. Then for all* $x\in[a,b]$*, we have the identity:*

$$\int_a^b g(t)\,dt=\sum_{k=0}^{n-1}\frac{1}{(k+1)!}\left[(b-x)^{k+1}+(-1)^k(x-a)^{k+1}\right]g^{(k)}(x)+(-1)^n\int_a^b K_n(x,t)\,g^{(n)}(t)\,dt, \tag{1.1}$$

where the kernel $K_n:[a,b]^2\to\mathbb{R}$ *is given by*

$$K_n(x,t):=\begin{cases}\frac{(t-a)^n}{n!}, & a\le t\le x\le b\\ \frac{(t-b)^n}{n!}, & a\le x<t\le b.\end{cases} \tag{1.2}$$

In particular, if $x=\frac{a+b}{2}$*, then*

$$\int_a^b g(t)\,dt=\sum_{k=0}^{n-1}\frac{1}{2^{k+1}}\left[\frac{1+(-1)^k}{(k+1)!}\right](b-a)^{k+1}g^{(k)}\left(\frac{a+b}{2}\right)+(-1)^n\int_a^b M_n(t)\,g^{(n)}(t)\,dt, \tag{1.3}$$

where

$$M_n(t):=\begin{cases}\frac{(t-a)^n}{n!}, & a\le t\le\frac{a+b}{2}\\ \frac{(t-b)^n}{n!}, & \frac{a+b}{2}<t\le b.\end{cases} \tag{1.4}$$

Another integral identity generalising the trapezoid rule is embodied in the following lemma (see [**17**] or [**14**]).

LEMMA 1.2. *Let* $g:[a,b]\to\mathbb{R}$ *be as in Lemma 1.1. Then for all* $x\in[a,b]$*, we have the representation*

$$\int_a^b g(t)\,dt=\sum_{k=0}^{n-1}\frac{1}{(k+1)!}\left[(x-a)^{k+1}g^{(k)}(a)+(-1)^k(b-x)^{k+1}g^{(k)}(b)\right]+\frac{1}{n!}\int_a^b(x-t)^n g^{(n)}(t)\,dt, \tag{1.5}$$

In particular, if $x = \frac{a+b}{2}$, then

$$\int_a^b g(t)\,dt = \sum_{k=0}^{n-1} \frac{1}{2^{k+1}(k+1)!}(b-a)^{k+1}\left[g^{(k)}(a) + (-1)^k g^{(k)}(b)\right] + \frac{(-1)^n}{n!}\int_a^b \left(t - \frac{a+b}{2}\right)^n g^{(n)}(t)\,dt. \tag{1.6}$$

Let us consider $x = (1-\lambda)a + \lambda b$, $\lambda \in [0,1]$, then from (1.1) we obtain

$$\begin{aligned}\int_a^b g(t)\,dt &= \sum_{k=0}^{n-1} \frac{1}{(k+1)!}\left[(1-\lambda)^{k+1} + (-1)^k \lambda^{k+1}\right](b-a)^{k+1} g^{(k)}((1-\lambda)a + \lambda b) \\ &+ (-1)^n \int_a^b K_n((1-\lambda)a + \lambda b, t)\, g^{(n)}(t)\,dt,\end{aligned} \tag{1.7}$$

and from (1.5) we obtain

$$\begin{aligned}\int_a^b g(t)\,dt &= \sum_{k=0}^{n-1} \frac{1}{(k+1)!}\left[\lambda^{k+1} g^{(k)}(a) + (-1)^k (1-\lambda)^{k+1} g^{(k)}(b)\right](b-a)^{k+1} \\ &+ \frac{1}{n!}\int_a^b [(1-\lambda)a + \lambda b - t]^n g^{(n)}(t)\,dt.\end{aligned} \tag{1.8}$$

We are now able to state and prove the following representation result for the Csiszár f-divergence.

THEOREM 1.3 (Barnett et al., 2002 [**5**]). *Let $f : \mathbb{R} \to \mathbb{R}$ be a function such that $f^{(n)}$ is absolutely continuous on any $[a,b] \subset \mathbb{R}$. If $p, q \in \mathbb{P}^m$, then*

$$\begin{aligned}I_f(p,q) &= f(1) + \sum_{k=0}^{n-1} \frac{1}{(k+1)!}\left[(1-\lambda)^{k+1} + (-1)^k \lambda^{k+1}\right] \\ &\times I_{(\cdot-1)^{k+1} f^{(k+1)}[(1-\lambda)+\lambda\cdot]}(p,q) + (-1)^n \sum_{i=1}^m q_i \\ &\times \left(\int_1^{\frac{p_i}{q_i}} K_n\left[\frac{(1-\lambda)q_i + \lambda p_i}{q_i}, t\right] f^{(n+1)}(t)\,dt\right), \ \lambda \in [0,1]\end{aligned} \tag{1.9}$$

and

$$\begin{aligned}I_f(p,q) &= f(1) + \sum_{k=0}^{n-1} \frac{\lambda^{k+1}}{(k+1)!} f^{(k+1)}(1)\, D_k(p,q) \\ &+ \sum_{k=0}^{n-1} \frac{(-1)^k (1-\lambda)^{k+1}}{(k+1)!} I_{(\cdot-1)^{k+1} f^{(k+1)}(\cdot)}(p,q) + \frac{1}{n!}\sum_{i=1}^m q_i^{-n+1} \\ &\times \left(\int_1^{\frac{p_i}{q_i}} [\lambda p_i + [(1-\lambda) - t]\, q_i]^n f^{(n+1)}(t)\,dt\right),\end{aligned} \tag{1.10}$$

where

$$D_k(p,q)=\sum_{i=1}^{m}(p_i-q_i)^k q_i^{-k+1}.$$

PROOF. If we apply the identity (1.7) for f', we get

$$f(b)=f(a)+\sum_{k=0}^{n-1}\frac{1}{(k+1)!}\left[(1-\lambda)^{k+1}+(-1)^k\lambda^{k+1}\right]\times(b-a)^{k+1}f^{(k+1)}((1-\lambda)a+\lambda b)+(-1)^n\int_a^b K_n[(1-\lambda)a+\lambda b,t]f^{(n+1)}(t)\,dt. \tag{1.11}$$

If in (1.11) we choose $b=\frac{p_i}{q_i}$, $i\in\{1,...,m\}$ and $a=1$, then we get

$$f\left(\frac{p_i}{q_i}\right)=f(1)+\sum_{k=0}^{n-1}\frac{1}{(k+1)!}\left[(1-\lambda)^{k+1}+(-1)^k\lambda^{k+1}\right]\times\frac{(p_i-q_i)^{k+1}}{[q_i]^{k+1}}\cdot f^{(k+1)}\left[\frac{(1-\lambda)q_i+\lambda p_i}{q_i}\right]+(-1)^n\int_1^{\frac{p_i}{q_i}}K_n\left[\frac{(1-\lambda)q_i+\lambda p_i}{q_i},t\right]f^{(n+1)}(t)\,dt \tag{1.12}$$

for all $i\in\{1,...,m\}$.

If we multiply (1.12) by $q_i\geq 0$ $(i\in\{1,...,m\})$, sum over $i\in\{1,...,m\}$ and take into account that $\sum_{i=1}^m q_i=1$, then we get the representation (1.9).

If we apply the identity (1.8) for f', we get

$$f(b)=f(a)+\sum_{k=0}^{n-1}\frac{1}{(k+1)!}\left[\lambda^{k+1}f^{(k+1)}(a)+(-1)^k(1-\lambda)^{k+1}f^{(k+1)}(b)\right](b-a)^{k+1}+\frac{1}{n!}\int_a^b[(1-\lambda)a+\lambda b-t]^n f^{(n+1)}(t)\,dt. \tag{1.13}$$

If in (1.13) we choose $b=\frac{p_i}{q_i}$, $i\in\{1,...,m\}$ and $a=1$, we get

$$f\left(\frac{p_i}{q_i}\right)=f(1)+\sum_{k=0}^{n-1}\frac{\lambda^{k+1}}{(k+1)!}f^{(k+1)}(1)\left(\frac{p_i}{q_i}-1\right)^{k+1}+\sum_{k=0}^{n-1}\frac{(-1)^k(1-\lambda)^{k+1}}{(k+1)!}f^{(k+1)}\left(\frac{p_i}{q_i}\right)\left(\frac{p_i}{q_i}-1\right)^{k+1}+\frac{1}{n!}\int_1^{\frac{p_i}{q_i}}\left[\frac{(1-\lambda)q_i+\lambda p_i}{q_i}-t\right]^n f^{(n+1)}(t)\,dt, \tag{1.14}$$

for all $i\in\{1,...,m\}$.

If we multiply (1.14) by $q_i\geq 0$ $(i\in\{1,...,m\})$, sum over $i\in\{1,...,m\}$ and take into account that $\sum_{i=1}^m q_i=1$, we get the representation (1.10). ∎

REMARK 1.1. If in (1.9) we choose $\lambda = 0$ or, $\lambda = 1$ or, $\lambda = \frac{1}{2}$, we get, respectively

$$I_f(p,q) = f(1) + \sum_{k=0}^{n-1} \frac{1}{(k+1)!} f^{(k+1)}(1) D_k(p,q) + (-1)^n \sum_{i=1}^{m} q_i \left(\int_1^{\frac{p_i}{q_i}} K_n(1,t) f^{(n+1)}(t)\, dt \right), \tag{1.15}$$

$$I_f(p,q) = f(1) + \sum_{k=0}^{n-1} \frac{(-1)^k}{(k+1)!} I_{(\cdot-1)^{k+1} f^{(k+1)}(\cdot)}(p,q) + (-1)^n \sum_{i=1}^{m} q_i \left(\int_1^{\frac{p_i}{q_i}} K_n\left(\frac{p_i}{q_i}, t\right) f^{(n+1)}(t)\, dt \right) \tag{1.16}$$

and

$$I_f(p,q) = f(1) + \sum_{k=0}^{n-1} \left[\frac{1 + (-1)^k}{2^{k+1}(k+1)!} \right] I_{(\cdot-1)^{k+1} f^{(k+1)}\left(\frac{1+\cdot}{2}\right)}(p,q) + (-1)^n \sum_{i=1}^{m} q_i \left(\int_1^{\frac{p_i}{q_i}} K_n\left(\frac{q_i + p_i}{2q_i}, t\right) f^{(n+1)}(t)\, dt \right). \tag{1.17}$$

REMARK 1.2. If in (1.10) we choose $\lambda = 0$, or $\lambda = 1$ or, $\lambda = \frac{1}{2}$, we get, respectively

$$I_f(p,q) = f(1) + \sum_{k=0}^{n-1} \frac{(-1)^k}{(k+1)!} I_{(\cdot-1)^{k+1} f^{(k+1)}(\cdot)}(p,q) + \frac{1}{n!} \sum_{i=1}^{m} q_i \left(\int_1^{\frac{p_i}{q_i}} (1-t)^n f^{(n+1)}(t)\, dt \right), \tag{1.18}$$

$$I_f(p,q) = f(1) + \sum_{k=0}^{n-1} \frac{f^{(k+1)}(1)}{(k+1)!} D_k(p,q) + \frac{1}{n!} \sum_{i=1}^{m} q_i^{-n+1} \left(\int_1^{\frac{p_i}{q_i}} (p_i - tq_i)^n f^{(n+1)}(t)\, dt \right) \tag{1.19}$$

and

$$I_f(p,q) = f(1) + \sum_{k=0}^{n-1} \frac{f^{(k+1)}(1)}{2^{k+1}(k+1)!} D_k(p,q) + \sum_{k=0}^{n-1} \frac{(-1)^k}{2^{k+1}(k+1)!} I_{(\cdot-1)^{k+1} f^{(k+1)}(\cdot)}(p,q) + \frac{1}{n!} \sum_{i=1}^{m} q_i^{-n+1} \left(\int_1^{\frac{p_i}{q_i}} \left[\frac{1}{2} p_i + \left(\frac{1}{2} - t \right) q_i \right]^n f^{(n+1)}(t)\, dt \right). \tag{1.20}$$

2. BOUNDS FOR THE REMAINDER

In this section we point out some bounds for the remainders in the representations (1.9) and (1.10), i.e.,

$$R_f(p,q) := (-1)^n \sum_{i=1}^{m} q_i \left(\int_1^{\frac{p_i}{q_i}} K_n \left[\frac{(1-\lambda) q_i + \lambda p_i}{q_i}, t \right] f^{(n+1)}(t)\, dt \right) \tag{2.1}$$

and

$$\tilde{R}_f(p,q) := \frac{1}{n!} \sum_{i=1}^{m} q_i^{-n+1} \int_1^{\frac{p_i}{q_i}} [\lambda p_i + [(1-\lambda) - t] q_i]^n f^{(n+1)}(t)\, dt \tag{2.2}$$

where $p, q \in \mathbb{P}^m$, $\lambda \in [0,1]$ and $K_n(\cdot,\cdot)$ is the kernel defined in equation (1.2).

For $a, b \in \mathbb{R}$, let us denote

$$\|f\|_{[a,b],p} := \left| \int_a^b |f(t)|^p\, dt \right|^{\frac{1}{p}}, \quad p \geq 1$$

and

$$\|f\|_{[a,b],\infty} := ess \sup_{\substack{t\in[a,b] \\ (t\in[b,a])}} |f(t)|.$$

In order to obtain bounds on $R_f(p,q)$ as given in (2.1), we need to consider integrals of the form

$$I_1(z) := \int_1^z K_n[(1-\lambda)\cdot 1 + \lambda z, t]\, f^{(n+1)}(t)\, dt, \quad z \in (0,\infty).$$

Thus

$$\begin{aligned}
|I_1(z)| &\leq \left| \int_1^z |K_n[(1-\lambda)\cdot 1 + \lambda z, t]| \left| f^{(n+1)}(t) \right| dt \right| \\
&\leq \left\| f^{(n+1)} \right\|_{[1,z],\infty} \left| \int_1^z |K_n((1-\lambda)\cdot 1 + \lambda z, t)|\, dt \right| \\
&= \frac{1}{n!} \left\| f^{(n+1)} \right\|_{[1,z],\infty} \left| \int_1^{(1-\lambda)\cdot 1+\lambda z} |t-1|^n\, dt + \int_{(1-\lambda)\cdot 1+\lambda z}^{z} |t-z|^n\, dt \right| \\
&= \frac{1}{n!} \left[\frac{|(1-\lambda) + \lambda z - 1|^{n+1} + |z - (1-\lambda)\cdot 1 - \lambda z|^{n+1}}{n+1} \right] \left\| f^{(n+1)} \right\|_{[1,z],\infty} \\
&= \frac{1}{n!} \left[\frac{\lambda^{n+1} |z-1|^{n+1} + (1-\lambda)^{n+1} |z-1|^{n+1}}{n+1} \right] \left\| f^{(n+1)} \right\|_{[1,z],\infty} \\
&= \frac{|z-1|^{n+1}}{(n+1)!} \left[\lambda^{n+1} + (1-\lambda)^{n+1} \right] \left\| f^{(n+1)} \right\|_{[1,z],\infty}.
\end{aligned}$$

Using Hölder's inequality, we may write for $\alpha > 1$, $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, that:

$$|I_1(z)| \leq \left\| f^{(n+1)} \right\|_{[1,z],\beta} \left| \int_1^z |K_n((1-\lambda)\cdot 1 + \lambda z, t)|^\alpha\, dt \right|^{\frac{1}{\alpha}}.$$

However,

$$\begin{aligned}
&\left|\int_1^z |K_n((1-\lambda)\cdot 1+\lambda z,t)|^\alpha dt\right|^{\frac{1}{\alpha}}\\
&=\frac{1}{n!}\left|\int_1^{(1-\lambda)\cdot 1+\lambda z}|t-1|^{\alpha n}dt+\int_{(1-\lambda)\cdot 1+\lambda z}^{z}|t-z|^{\alpha n}dt\right|^{\frac{1}{\alpha}}\\
&=\frac{1}{n!}\left[\frac{|(1-\lambda)+\lambda z-1|^{\alpha n+1}+|z-(1-\lambda)\cdot 1-\lambda z|^{\alpha n+1}}{\alpha n+1}\right]^{\frac{1}{\alpha}}\\
&=\frac{1}{n!}\left[\frac{\lambda^{\alpha n+1}|z-1|^{\alpha n+1}+(1-\lambda)^{\alpha n+1}|z-1|^{\alpha n+1}}{\alpha n+1}\right]^{\frac{1}{\alpha}}\\
&=\frac{|z-1|^{n+\frac{1}{\alpha}}}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\left[\lambda^{\alpha n+1}+(1-\lambda)^{\alpha n+1}\right]^{\frac{1}{\alpha}}
\end{aligned}$$

and then:

$$|I_1(z)|\le\left\|f^{(n+1)}\right\|_{[1,z],\beta}\frac{|z-1|^{n+\frac{1}{\alpha}}}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\left[\lambda^{\alpha n+1}+(1-\lambda)^{\alpha n+1}\right]^{\frac{1}{\alpha}}.$$

Finally, we observe that

$$\begin{aligned}
&\sup_{t\in[1,z]}|K_n((1-\lambda)\cdot 1+\lambda z,t)|\\
&=\frac{1}{n!}\max\left\{((1-\lambda)+\lambda z-1)^n+(z-(1-\lambda)\cdot 1-\lambda z)^n\right\}\\
&=\frac{1}{n!}(z-1)^n(\max\{\lambda,1-\lambda\})^n=\frac{1}{n!}|z-1|^n\left[\frac{1}{2}+\left|\lambda-\frac{1}{2}\right|\right]^n
\end{aligned}$$

and then

$$|I_1(z)|\le\frac{1}{n!}|z-1|^n\left[\frac{1}{2}+\left|\lambda-\frac{1}{2}\right|\right]^n\left\|f^{(n+1)}\right\|_{[1,z],1}.$$

Using the above inequalities, we may state the following result

$$|I_1(z)|\le\left\{\begin{array}{l}
\frac{|z-1|^{n+1}}{(n+1)!}\left[\lambda^{n+1}+(1-\lambda)^{n+1}\right]\left\|f^{(n+1)}\right\|_{[1,z],\infty}\\[2ex]
\frac{|z-1|^{n+\frac{1}{\alpha}}}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\left[\lambda^{\alpha n+1}+(1-\lambda)^{\alpha n+1}\right]^{\frac{1}{\alpha}}\left\|f^{(n+1)}\right\|_{[1,z],\beta}\\
\qquad\qquad\alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1\\[2ex]
\frac{1}{n!}|z-1|^n\left[\frac{1}{2}+\left|\lambda-\frac{1}{2}\right|\right]^n\left\|f^{(n+1)}\right\|_{[1,z],1}
\end{array}\right\}=:\kappa(z,n) \tag{2.3}$$

for all $z>0$, $n\in\mathbb{N}$.

We are now able to state the following theorem pertaining to the remainder $R_f(p,q)$.

THEOREM 2.1 (Barnett et al., 2002 [**5**]). *Assume that the function f is as in Theorem 1.3. If $p, q \in \mathbb{P}^m$, then we have the inequality*

$$\text{(2.4)} \quad |R_f(p,q)| \leq A := \begin{cases} \frac{1}{(n+1)!}\left[\lambda^{n+1} + (1-\lambda)^{n+1}\right] \sum_{i=1}^m q_i^{-n} |p_i - q_i|^{n+1} \left\| f^{(n+1)} \right\|_{\left[1, \frac{p_i}{q_i}\right], \infty} \\ \\ \frac{1}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\left[\lambda^{\alpha n+1} + (1-\lambda)^{\alpha n+1}\right]^{\frac{1}{\alpha}} \\ \times \sum_{i=1}^m q_i^{-n-\frac{1}{\alpha}+1} |p_i - q_i|^{n+\frac{1}{\alpha}} \left\| f^{(n+1)} \right\|_{\left[1, \frac{p_i}{q_i}\right], \beta} \\ \text{where } \alpha > 1, \ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \\ \frac{1}{n!}\left[\frac{1}{2} + \left|\lambda - \frac{1}{2}\right|\right]^n \sum_{i=1}^m q_i^{-n+1} |p_i - q_i|^n \left\| f^{(n+1)} \right\|_{\left[1, \frac{p_i}{q_i}\right], 1}. \end{cases}$$

Moreover, if we assume that $0 \leq r \leq \frac{p_i}{q_i} \leq R < \infty$, $i \in \{1, ..., n\}$, then the second term in (2.4) can be upper bounded by

$$\text{(2.5)} \qquad B := \begin{cases} \frac{1}{(n+1)!}\left[\lambda^{n+1} + (1-\lambda)^{n+1}\right] \left\| f^{(n+1)} \right\|_{[r,R], \infty} \\ \times \sum_{i=1}^m q_i^{-n} |p_i - q_i|^{n+1} \\ \\ \frac{1}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\left[\lambda^{\alpha n+1} + (1-\lambda)^{\alpha n+1}\right]^{\frac{1}{\alpha}} \left\| f^{(n+1)} \right\|_{[r,R], \beta} \\ \times \sum_{i=1}^m q_i^{-n-\frac{1}{\alpha}+1} |p_i - q_i|^{n+\frac{1}{\alpha}} \quad \text{if} \quad \alpha > 1, \ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \\ \frac{1}{n!}\left[\frac{1}{2} + \left|\lambda - \frac{1}{2}\right|\right]^n \left\| f^{(n+1)} \right\|_{[r,R], 1} \sum_{i=1}^m q_i^{-n+1} |p_i - q_i|^n \end{cases}$$

$$\leq C := \begin{cases} \frac{1}{(n+1)!}\left[\lambda^{n+1} + (1-\lambda)^{n+1}\right] \left\| f^{(n+1)} \right\|_{[r,R], \infty} (R-r)^{n+1} \\ \\ \frac{1}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\left[\lambda^{\alpha n+1} + (1-\lambda)^{\alpha n+1}\right]^{\frac{1}{\alpha}} \\ \times \left\| f^{(n+1)} \right\|_{[r,R], \beta} (R-r)^{n+\frac{1}{\alpha}} \quad \text{if} \ \alpha > 1, \ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \\ \frac{1}{n!}\left[\frac{1}{2} + \left|\lambda - \frac{1}{2}\right|\right]^n \left\| f^{(n+1)} \right\|_{[r,R], 1} (R-r)^n. \end{cases}$$

The proof of (2.4) follows by the inequality (2.3) choosing $z = \frac{p_i}{q_i}$ and suming over $i \in \{1, ..., m\}$.

The proof of (2.5) follows by the fact that $\left|\frac{p_i}{q_i} - 1\right| \leq R - r$ for all $i \in \{1, ..., n\}$.

We omit the details.

The following corollary may be useful in practical applications.

COROLLARY 2.2 (Barnett et al., 2002 [**5**]). *With the assumptions of Theorem 2.1, we have the inequality:*

$$\left|I_f(p,q)-f(1)-\sum_{k=0}^{n-1}\frac{f^{(k+1)}(1)}{(k+1)!}D_k(p,q)\right| \tag{2.6}$$

$$\leq \left\{\begin{array}{l} \frac{1}{(n+1)!}\sum_{i=1}^m q_i^{-n}\left|p_i-q_i\right|^{n+1}\left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],\infty} \\ \\ \frac{1}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\sum_{i=1}^m q_i^{-n-\frac{1}{\alpha}+1}\left|p_i-q_i\right|^{n+\frac{1}{\alpha}}\left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],\beta} \\ \qquad \textit{where } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \\ \frac{1}{n!}\sum_{i=1}^m q_i^{-n+1}\left|p_i-q_i\right|^{n}\left\|f^{(n+1)}\right\|_{\left[1,\frac{p_i}{q_i}\right],1} \end{array}\right\} =: M_1.$$

Furthermore, if we assume that $r\leq\frac{p_i}{q_i}\leq R<\infty$*,* $i\in\{1,...,n\}$*, then we have*

$$M_1 \leq \left\{\begin{array}{l} \frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_{[r,R],\infty}\sum_{i=1}^m q_i^{-n}\left|p_i-q_i\right|^{n+1} \\ \\ \frac{1}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\left\|f^{(n+1)}\right\|_{[r,R],\beta}\sum_{i=1}^m q_i^{-n-\frac{1}{\alpha}+1}\left|p_i-q_i\right|^{n+\frac{1}{\alpha}} \\ \qquad \textit{where } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[r,R],1}\sum_{i=1}^m q_i^{-n+1}\left|p_i-q_i\right|^{n} \end{array}\right\} =: M_2$$

$$\leq \left\{\begin{array}{l} \frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_{[r,R],\infty}(R-r)^{n+1} \\ \\ \frac{1}{n!(\alpha n+1)^{\frac{1}{\alpha}}}\left\|f^{(n+1)}\right\|_{[r,R],\beta}(R-r)^{n+\frac{1}{\alpha}} \\ \qquad \textit{where } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[r,R],1}(R-r)^{n} \end{array}\right\} =: M_3$$

and

$$\left|I_f(p,q)-f(1)-\sum_{k=0}^{n-1}\frac{(-1)^k}{(k+1)!}I_{(\cdot-1)^{k+1}f^{(k+1)}(\cdot)}(p,q)\right| \leq M_1\leq M_2\leq M_3$$

and if $0\leq r\leq\frac{p_i}{q_i}\leq R<\infty$*,* $i\in\{1,...,n\}$*, then*

$$\left|I_f(p,q)-f(1)-\sum_{k=0}^{n-1}\left[\frac{1+(-1)^k}{2^{k+1}(k+1)!}\right]I_{(\cdot-1)^{k+1}f^{(k+1)}\left(\frac{1+\cdot}{2}\right)}(p,q)\right| \leq \frac{1}{2^n}M_1\leq\frac{1}{2^n}M_2\leq\frac{1}{2^n}M_3.$$

Now, to obtain the bound on $\tilde{R}_f(p,q)$ as defined in (2.2), consider the integral

$$I_2(z):=\frac{1}{n!}\int_1^z((1-\lambda)\cdot 1+\lambda z-t)^n f^{(n+1)}(t)\,dt,$$

from which we have

$$
\begin{aligned}
|I_2(z)| &\leq \left\|f^{(n+1)}\right\|_{[1,z],\infty} \frac{1}{n!}\left|\int_1^z |(1-\lambda)\cdot 1+\lambda z-t|^n\right| dt \\
&= \frac{1}{n!}\left|\int_1^{(1-\lambda)\cdot 1+\lambda z} |(1-\lambda)\cdot 1+\lambda z-t|^n\, dt\right. \\
&\left.+ \int_{(1-\lambda)\cdot 1+\lambda z}^{z} |(1-\lambda)\cdot 1+\lambda z-t|^n\, dt\right| \left\|f^{(n+1)}\right\|_{[1,z],\infty} \\
&= \frac{1}{n!}\cdot\left[\frac{|(1-\lambda)\cdot 1+\lambda z-1|^{n+1}+|(1-\lambda)\cdot 1+\lambda z-z|^{n+1}}{n+1}\right] \\
&\times \left\|f^{(n+1)}\right\|_{[1,z],\infty} \\
&= \frac{(z-1)^{n+1}}{(n+1)!}\cdot\left[\lambda^{n+1}+(1-\lambda)^{n+1}\right]\cdot\left\|f^{(n+1)}\right\|_{[1,z],\infty}.
\end{aligned}
$$

Using Hölder's inequality, we may write, for $\alpha > 1$, $\frac{1}{\alpha}+\frac{1}{\beta}=1$, that

$$
\begin{aligned}
|I_2(z)| &\leq \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[1,z],\beta}\left|\int_1^z |(1-\lambda)\cdot 1+\lambda z-t|^{n\alpha}\, dt\right|^{\frac{1}{\alpha}} \\
&= \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[1,z],\beta}\left|\int_1^{(1-\lambda)\cdot 1+\lambda z} |(1-\lambda)\cdot 1+\lambda z-t|^{n\alpha}\, dt\right. \\
&\left.+ \int_{(1-\lambda)\cdot 1+\lambda z}^{z} |(1-\lambda)\cdot 1+\lambda z-t|^{n\alpha}\, dt\right|^{\frac{1}{\alpha}} \\
&= \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[1,z],\beta}\left[\frac{|(1-\lambda)+\lambda z-1|^{n\alpha+1}+|z-(1-\lambda)-\lambda z|^{n\alpha+1}}{n\alpha+1}\right]^{\frac{1}{\alpha}} \\
&= \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[1,z],\beta}\left[\frac{\lambda^{\alpha n+1}|z-1|^{\alpha n+1}+(1-\lambda)^{\alpha n+1}|z-1|^{\alpha n+1}}{\alpha n+1}\right]^{\frac{1}{\alpha}} \\
&= \frac{|z-1|^{n+\frac{1}{\alpha}}}{n!\,(\alpha n+1)^{\frac{1}{\alpha}}}\cdot\left[\lambda^{\alpha n+1}+(1-\lambda)^{\alpha n+1}\right]^{\frac{1}{\alpha}}\cdot\left\|f^{(n+1)}\right\|_{[1,z],\beta}.
\end{aligned}
$$

Finally, we observe that

$$
\begin{aligned}
|I_2(z)| &\leq \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[1,z],1}\sup_{t\in[1,z]}|(1-\lambda)\cdot 1+\lambda z-t|^n \\
&= \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[1,z],1}\max\left\{|(1-\lambda)+\lambda z-1|^n+|z-(1-\lambda)\cdot 1-\lambda z|^n\right\} \\
&= \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[1,z],1}(z-1)^n\left(\max\{\lambda,1-\lambda\}\right)^n \\
&= \frac{1}{n!}\left\|f^{(n+1)}\right\|_{[1,z],1}|z-1|^n\left[\frac{1}{2}+\left|\lambda-\frac{1}{2}\right|\right]^n.
\end{aligned}
$$

Using the above inequalities, we may state that

$$
|I_2(z)| \leq \kappa(n,z), \tag{2.7}
$$

where $\kappa(n,z)$ is defined in (2.3). That is, the bounds for $R_f(p,q)$ and $\tilde{R}_f(p,q)$ are the same.

We may now state the following theorem concerning a bound for the remainder $\tilde{R}_f(p,q)$.

THEOREM 2.3 (Barnett et al., 2002 [**5**]). *Assume that the function f is as in Theorem 1.3. If $p,q \in \mathbb{P}^m$, then we have the inequality:*

$$\left|\tilde{R}_f(p,q)\right| \le A, \tag{2.8}$$

where A is given in (2.4).
Moreover, if we assume that $0 \le r \le \frac{p_i}{q_i} \le R < \infty$, $i \in \{1,...,n\}$, then

$$A \le B \le C, \tag{2.9}$$

with B and C being as defined in (2.5).

The following corollary may be useful in practical applications.

COROLLARY 2.4 (Barnett et al., 2002 [**5**]). *With the above assumptions, we have*

$$\left|I_f(p,q) - f(1) - \sum_{k=0}^{n-1} \frac{(-1)^k}{(k+1)!} I_{(\cdot-1)^{k+1} f^{(k+1)}(\cdot)}(p,q)\right| \le M_1 \le M_2 \le M_3, \tag{2.10}$$

$$\left|I_f(p,q) - f(1) - \sum_{k=0}^{n-1} \frac{f^{(k+1)}(1)}{(k+1)!} D_k(p,q)\right| \le M_1 \le M_2 \le M_3 \tag{2.11}$$

and

$$\left|I_f(p,q) - f(1) - \sum_{k=0}^{n-1} \frac{f^{(k+1)}(1)}{2^{k+1}(k+1)!} D_k(p,q) - \sum_{k=0}^{n-1} \frac{(-1)^k}{2^{k+1}(k+1)!} I_{(\cdot-1)^{k+1} f^{(k+1)}(\cdot)}(p,q)\right| \le \frac{1}{2^n} M_1 \le \frac{1}{2^n} M_2 \le \frac{1}{2^n} M_3 \tag{2.12}$$

for $r \le \frac{p_i}{q_i} \le R$, $i \in \{1,...,n\}$, where M_i $\left(i = \overline{1,3}\right)$ are as defined in Corollary 2.2.

REMARK 2.1. If in all the above results we choose f to be a particular function generating the classical divergences listed in the introduction), then we can obtain many interesting approximations for the above distances. We omit the details.

3. REPRESENTATION OF f-DIVERGENCE VIA A GENERAL MONTGOMERY IDENTITY

In [**16**], the authors have pointed out the following integral identity generalising the Montgomery identity.

LEMMA 3.1. *Let $f : [a,b] \to \mathbb{R}$ be a mapping such that $f^{(n-1)}$ is absolutely continuous on $[a,b]$. Then for all $x \in [a,b]$, we have the identity:*

$$\int_a^b f(t)\,dt = \sum_{k=0}^{n-1} \left[\frac{(b-x)^{k+1} + (-1)^k (x-a)^{k+1}}{(k+1)!}\right] f^{(k)}(x) + (-1)^n \int_a^b K_n(x,t) f^{(n)}(t)\,dt \tag{3.1}$$

where the kernel $K_n : [a,b]^2 \to \mathbb{R}$ *is given by*

$$K_n(x,t) := \begin{cases} \frac{(t-a)^n}{n!}, & a \le t \le x \le b \\ \\ \frac{(t-b)^n}{n!}, & a \le x < t \le b \end{cases} \tag{3.2}$$

and n *is a natural number,* $n \ge 1$.

In what follows, we need the identity (3.2) in the following equivalent form [**16**]:

$$\begin{aligned} f(z) &= \frac{1}{b-a}\int_a^b f(t)\,dt \\ &\quad - \frac{1}{b-a}\sum_{k=1}^{n-1}\frac{1}{(k+1)!}\left[(b-z)^{k+1} + (-1)^k (z-a)^{k+1}\right] f^{(k)}(z) \\ &\quad + \frac{(-1)^{n+1}}{(b-a)\,n!}\left[\int_a^z (t-a)^n f^{(n)}(t)\,dt + \int_z^b (t-b)^n f^{(n)}(t)\,dt\right] \end{aligned} \tag{3.3}$$

for all $z \in [a,b]$.

Note that for $n = 1$, the sum $\sum_{k=1}^{n-1}$ is empty and we obtain the *Montgomery identity* (see for example [**35**])

$$\begin{aligned} f(z) = \frac{1}{b-a}\int_a^b f(t)\,dt + \frac{1}{b-a}\left[\int_a^z (t-a) f^{(1)}(t)\,dt \right.\\ \left. + \int_z^b (t-b) f^{(1)}(t)\,dt\right],\ x \in [a,b]. \end{aligned} \tag{3.4}$$

In what follows, we assume that the probability distributions $p, q \in \mathbb{P}^m$ satisfy the standing condition:

$$0 \le r \le \frac{p_i}{q_i} \le R < \infty, \text{ for } i \in \{1, ..., m\}. \tag{3.5}$$

Obviously $r \le 1 \le R$.

The following representation of Csiszár $f-$divergence holds.

THEOREM 3.2 (Barnett et al., 2002 [**7**]). *Let* $f : [r,R] \to \mathbb{R}$*, where* r, R *are as above and* $f^{(n-1)}$ *is absolutely continuous on* $[r,R]$*. Then for all* $p, q \in \mathbb{P}^m$ *satisfying (3.5), we have the representation:*

$$\begin{aligned} &I_f(p,q) \\ &= \frac{1}{R-r}\int_r^R f(t)\,dt - \frac{1}{R-r}\sum_{k=1}^{n-1}\frac{1}{(k+1)!} I_{(R-\cdot)^{k+1} f^{(k)}(\cdot)}(p,q) \\ &+ \frac{1}{R-r}\sum_{k=1}^{n-1}\frac{(-1)^{k+1}}{(k+1)!} I_{(\cdot-r)^{k+1} f^{(k)}(\cdot)}(p,q) \\ &+ \frac{(-1)^{n+1}}{(R-r)\,n!}\sum_{i=1}^m q_i \left(\int_r^{\frac{p_i}{q_i}} (t-r)^n f^{(n)}(t)\,dt\right) \\ &+ \frac{(-1)^{n+1}}{(R-r)\,n!}\sum_{i=1}^m q_i \left(\int_{\frac{p_i}{q_i}}^R (t-R)^n f^{(n)}(t)\,dt\right). \end{aligned} \tag{3.6}$$

PROOF. Using (3.3) for $z = \frac{p_i}{q_i}$, $i \in \{1, ..., m\}$ and $a = r$, $b = R$, we may write

$$
\begin{aligned}
(3.7) \qquad & f\left(\frac{p_i}{q_i}\right) \\
&= \frac{1}{R-r}\int_r^R f(t)\,dt - \frac{1}{R-r}\sum_{k=1}^{n-1}\frac{1}{(k+1)!}\left[\left(R-\frac{p_i}{q_i}\right)^{k+1}\right. \\
&\left. + (-1)^k\left(\frac{p_i}{q_i}-r\right)^{k+1}\right] f^{(k)}\left(\frac{p_i}{q_i}\right) \\
&+ \frac{(-1)^{n+1}}{(R-r)\,n!}\left[\int_r^{\frac{p_i}{q_i}}(t-r)^n f^{(n)}(t)\,dt + \int_{\frac{p_i}{q_i}}^R (t-R)^n f^{(n)}(t)\,dt\right]
\end{aligned}
$$

for all $i \in \{1, ..., m\}$.

If we multiply (3.7) by $q_i \geq 0$, sum over $i \in \{1, ...m\}$ and take into account that $\sum_{i=1}^m q_i = 1$, we get the desired identity (3.6). ∎

REMARK 3.1. If $n = 1$, then we have the representation:

$$
\begin{aligned}
(3.8) \qquad I_f(p,q) &= \frac{1}{R-r}\int_r^R f(t)\,dt \\
&+ \frac{1}{R-r}\sum_{i=1}^m q_i\left(\int_r^{\frac{p_i}{q_i}}(t-r)\,f'(t)\,dt\right) \\
&+ \frac{1}{R-r}\sum_{i=1}^m q_i\left(\int_{\frac{p_i}{q_i}}^R (t-R)\,f'(t)\,dt\right)
\end{aligned}
$$

if $n = 2$, then we have the representation:

$$
\begin{aligned}
(3.9) \qquad I_f(p,q) &= \frac{1}{R-r}\int_r^R f(t)\,dt + \frac{1}{2(R-r)}I_{(R-\cdot)^2 f^{(1)}(\cdot)}(p,q) \\
&+ \frac{1}{R-r}I_{(\cdot-r)^2 f'(\cdot)}(p,q) \\
&- \frac{1}{2(R-r)}\sum_{i=1}^m q_i\left(\int_r^{\frac{p_i}{q_i}}(t-r)^2 f^{(2)}(t)\,dt\right) \\
&- \frac{1}{2(R-r)}\sum_{i=1}^m q_i\left(\int_{\frac{p_i}{q_i}}^R (t-R)^2 f^{(2)}(t)\,dt\right).
\end{aligned}
$$

4. BOUNDS FOR THE REMAINDER

In formula (3.6), we consider the remainder $R_f(p,q)$ given by

$$
\begin{aligned}
R_f(p,q) := \frac{(-1)^{n+1}}{(R-r)\,n!}&\left[\sum_{i=1}^m q_i\left(\int_r^{\frac{p_i}{q_i}}(t-r)^n f^{(n)}(t)\,dt\right)\right. \\
&\left. + \sum_{i=1}^m q_i\left(\int_{\frac{p_i}{q_i}}^R (t-R)^n f^{(n)}(t)\,dt\right)\right].
\end{aligned}
$$

In this section, we are interested in obtaining some bounds for $R_f(p,q)$.

For this purpose, consider

$$I_1(z) = \int_r^z (t-r)^n f^{(n)}(t)\,dt$$

and

$$I_2(z) := \int_z^R (t-R)^n f^{(n)}(t)\,dt,$$

where $z \in [r, R]$.

For $a < b$, we also define the Lebesgue norm,

$$\|f\|_{[a,b],p} := \left[\int_a^b |f(t)|^p\,dt\right]^{\frac{1}{p}}, \quad p \geq 1$$

and

$$\|f\|_{[a,b],\infty} := ess \sup_{t\in[a,b]} |f(t)|.$$

Now, we observe that

$$|I_1(z)| \leq \int_r^z |t-r|^n \left|f^{(n)}(t)\right| dt \leq \left\|f^{(n)}\right\|_{[r,z],\infty} \frac{(z-r)^{n+1}}{n+1},$$

$$|I_1(z)| \leq \left\|f^{(n)}\right\|_{[r,z],\beta} \left(\int_r^z |t-r|^{\alpha n}\,dt\right)^{\frac{1}{\alpha}} = \left\|f^{(n)}\right\|_{[r,z],\beta} \left[\frac{(z-r)^{\alpha n+1}}{\alpha n+1}\right]^{\frac{1}{\alpha}}$$

$$= \left\|f^{(n)}\right\|_{[r,z],\beta} \frac{(z-r)^{n+\frac{1}{\alpha}}}{(\alpha n+1)^{\frac{1}{\alpha}}} \text{ if } \alpha > 1, \frac{1}{\alpha}+\frac{1}{\beta} = 1,$$

and, finally,

$$|I_1(z)| \leq \left\|f^{(n)}\right\|_{[r,z],1} \sup_{t\in[r,z]} |t-r|^n = \left\|f^{(n)}\right\|_{[r,z],1} (z-r)^n.$$

Consequently, we have

$$|I_1(z)| \leq \begin{cases} \frac{(z-r)^{n+1}}{n+1} \left\|f^{(n)}\right\|_{[r,z],\infty} \\ \\ \frac{(z-r)^{n+\frac{1}{\alpha}}}{(\alpha n+1)^{\frac{1}{\alpha}}} \left\|f^{(n)}\right\|_{[r,z],\beta}, \quad \alpha > 1, \ \frac{1}{\alpha}+\frac{1}{\beta} = 1; \\ \\ (z-r)^n \left\|f^{(n)}\right\|_{[r,z],1}. \end{cases} \tag{4.1}$$

In a similar fashion, we may point out that

$$|I_2(z)| \leq \begin{cases} \frac{(R-z)^{n+1}}{n+1} \left\|f^{(n)}\right\|_{[z,R],\infty} \\ \\ \frac{(R-z)^{n+\frac{1}{\alpha}}}{(\alpha n+1)^{\frac{1}{\alpha}}} \left\|f^{(n)}\right\|_{[z,R],\beta}, \quad \alpha > 1, \ \frac{1}{\alpha}+\frac{1}{\beta} = 1; \\ \\ (R-z)^n \left\|f^{(n)}\right\|_{[z,R],1}. \end{cases} \tag{4.2}$$

The following bound for the remainder $R_f(p,q)$ holds.

THEOREM 4.1 (Barnett et al., 2002 [**7**]). *Let f, r, R, p and q be as in Theorem 3.2. Then we have the inequality*

$$|R_f(p,q)| \leq \frac{1}{n!(R-r)} \tag{4.3}$$

$$\times \begin{cases} \frac{1}{n+1}\sum_{i=1}^{m} q_i \left[\left(R-\frac{p_i}{q_i}\right)^{n+1} \left\|f^{(n)}\right\|_{\left[\frac{p_i}{q_i},R\right],\infty} \right. \\ \qquad \left. + \left(\frac{p_i}{q_i}-r\right)^{n+1} \left\|f^{(n)}\right\|_{\left[r,\frac{p_i}{q_i}\right],\infty} \right] \\ \\ \frac{1}{(\alpha n+1)^{\frac{1}{\alpha}}}\sum_{i=1}^{m} q_i \left[\left(R-\frac{p_i}{q_i}\right)^{n+\frac{1}{\alpha}} \left\|f^{(n)}\right\|_{\left[\frac{p_i}{q_i},R\right],\beta} \right. \\ \qquad \left. + \left(\frac{p_i}{q_i}-r\right)^{n+\frac{1}{\alpha}} \left\|f^{(n)}\right\|_{\left[r,\frac{p_i}{q_i}\right],\beta} \right] \\ \textit{if } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \\ \sum_{i=1}^{m} q_i \left[\left(R-\frac{p_i}{q_i}\right)^{n} \left\|f^{(n)}\right\|_{\left[\frac{p_i}{q_i},R\right],1} \right. \\ \qquad \left. + \left(\frac{p_i}{q_i}-r\right)^{n} \left\|f^{(n)}\right\|_{\left[r,\frac{p_i}{q_i}\right],1} \right] \end{cases}$$

$$\le \frac{1}{n!\,(R-r)} \times \begin{cases} \frac{\left\|f^{(n)}\right\|_{[r,R],\infty}}{n+1}\sum_{i=1}^{m} q_i \left[\left(R-\frac{p_i}{q_i}\right)^{n+1} + \left(\frac{p_i}{q_i}-r\right)^{n+1} \right] \\ \\ \frac{\left\|f^{(n)}\right\|_{[r,R],\beta}}{(\alpha n+1)^{\frac{1}{\alpha}}}\sum_{i=1}^{m} q_i \left[\left(R-\frac{p_i}{q_i}\right)^{n+\frac{1}{\alpha}} + \left(\frac{p_i}{q_i}-r\right)^{n+\frac{1}{\alpha}} \right] \\ \\ \left\|f^{(n)}\right\|_{[r,R],1}\sum_{i=1}^{m} q_i \left[\left(R-\frac{p_i}{q_i}\right)^{n} + \left(\frac{p_i}{q_i}-r\right)^{n} \right] \end{cases}$$

$$\le \begin{cases} \frac{\left\|f^{(n)}\right\|_{[r,R],\infty}(R-r)^n}{(n+1)!} \\ \\ \frac{\left\|f^{(n)}\right\|_{[r,R],\beta}(R-r)^{n+\frac{1}{\alpha}-1}}{n!(\alpha n+1)^{\frac{1}{\alpha}}} \quad \textit{if } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \\ \frac{\left\|f^{(n)}\right\|_{[r,R],1}(R-r)^{n-1}}{n!} \end{cases}$$

PROOF. Using (4.1) and (4.2), we may write:

$$\begin{aligned}
|R_f(p,q)| &\leq \frac{1}{n!\,(R-r)}\left[\sum_{i=1}^{m} q_i\left|\int_r^{\frac{p_i}{q_i}} (t-r)^n f^{(n)}(t)\,dt\right|\right. \\
&\quad\left.+\sum_{i=1}^{m} q_i\left|\int_{\frac{p_i}{q_i}}^{R} (t-R)^n f^{(n)}(t)\,dt\right|\right] \\
&\leq \frac{1}{(R-r)\,n!}\times\begin{cases}
\frac{1}{n+1}\sum_{i=1}^{m} q_i\left(\frac{p_i}{q_i}-r\right)^{n+1}\left\|f^{(n)}\right\|_{\left[r,\frac{p_i}{q_i}\right],\infty} \\
\frac{1}{(\alpha n+1)^{\frac{1}{\alpha}}}\sum_{i=1}^{m} q_i\left(\frac{p_i}{q_i}-r\right)^{n+\frac{1}{\alpha}}\left\|f^{(n)}\right\|_{\left[r,\frac{p_i}{q_i}\right],\beta} \\
\text{if } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\
\sum_{i=1}^{m} q_i\left(\frac{p_i}{q_i}-r\right)^{n}\left\|f^{(n)}\right\|_{\left[r,\frac{p_i}{q_i}\right],1}
\end{cases} \\
&\quad+\frac{1}{(R-r)\,n!}\times\begin{cases}
\frac{1}{n+1}\sum_{i=1}^{m} q_i\left(R-\frac{p_i}{q_i}\right)^{n+1}\left\|f^{(n)}\right\|_{\left[\frac{p_i}{q_i},R\right],\infty} \\
\frac{1}{(\alpha n+1)^{\frac{1}{\alpha}}}\sum_{i=1}^{m} q_i\left(R-\frac{p_i}{q_i}\right)^{n+\frac{1}{\alpha}}\left\|f^{(n)}\right\|_{\left[\frac{p_i}{q_i},R\right],\beta} \\
\text{if } \alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\
\sum_{i=1}^{m} q_i\left(R-\frac{p_i}{q_i}\right)^{n}\left\|f^{(n)}\right\|_{\left[\frac{p_i}{q_i},R\right],1}
\end{cases} \\
&\leq \frac{1}{(R-r)\,n!}\times\begin{cases}
\frac{\left\|f^{(n)}\right\|_{[r,R],\infty}}{n+1}\sum_{i=1}^{m} q_i\left[\left(\frac{p_i}{q_i}-r\right)^{n+1}+\left(R-\frac{p_i}{q_i}\right)^{n+1}\right] \\
\frac{\left\|f^{(n)}\right\|_{[r,R],\beta}}{(\alpha n+1)^{\frac{1}{\alpha}}}\sum_{i=1}^{m} q_i\left[\left(\frac{p_i}{q_i}-r\right)^{n+\frac{1}{\alpha}}+\left(R-\frac{p_i}{q_i}\right)^{n+\frac{1}{\alpha}}\right] \\
\left\|f^{(n)}\right\|_{[r,R],1}\sum_{i=1}^{m} q_i\left[\left(\frac{p_i}{q_i}-r\right)^{n}+\left(R-\frac{p_i}{q_i}\right)^{n}\right]
\end{cases} \\
&\leq \frac{1}{(R-r)\,n!}\times\begin{cases}
\frac{\left\|f^{(n)}\right\|_{[r,R],\infty}}{n+1}(R-r)^{n+1} \\
\frac{\left\|f^{(n)}\right\|_{[r,R],\beta}(R-r)^{n+\frac{1}{\alpha}}}{(\alpha n+1)^{\frac{1}{\alpha}}} \\
\left\|f^{(n)}\right\|_{[r,R],1}(R-r)^{n}
\end{cases}
=\begin{cases}
\frac{\left\|f^{(n)}\right\|_{[r,R],\infty}(R-r)^{n}}{(n+1)!} \\
\frac{\left\|f^{(n)}\right\|_{[r,R],\beta}(R-r)^{n+\frac{1}{\alpha}-1}}{n!(\alpha n+1)^{\frac{1}{\alpha}}} \\
\frac{\left\|f^{(n)}\right\|_{[r,R],1}(R-r)^{n-1}}{n!}
\end{cases}
\end{aligned}$$

and the theorem is proved. ∎

REMARK 4.1. For $n = 1$, we obtain the estimate

$$|R_f(p,q)| \leq \frac{1}{R-r} \times \begin{cases} \frac{1}{2}\sum_{i=1}^m q_i \Big[\left(R - \frac{p_i}{q_i}\right)^2 \left\| f^{(1)} \right\|_{\left[\frac{p_i}{q_i},R\right],\infty} \\ \qquad + \left(\frac{p_i}{q_i} - r\right)^2 \left\| f^{(1)} \right\|_{\left[r,\frac{p_i}{q_i}\right],\infty} \Big] \\ \\ \frac{1}{(\alpha+1)^{\frac{1}{\alpha}}}\sum_{i=1}^m q_i \Big[\left(R - \frac{p_i}{q_i}\right)^{1+\frac{1}{\alpha}} \left\| f^{(1)} \right\|_{\left[\frac{p_i}{q_i},R\right],\beta} \\ \qquad + \left(\frac{p_i}{q_i} - r\right)^{1+\frac{1}{\alpha}} \left\| f^{(1)} \right\|_{\left[r,\frac{p_i}{q_i}\right],\beta} \Big] \\ \text{if } \alpha > 1,\ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \\ \sum_{i=1}^m q_i \Big[\left(R - \frac{p_i}{q_i}\right) \left\| f^{(1)} \right\|_{\left[\frac{p_i}{q_i},R\right],1} \\ \qquad + \left(\frac{p_i}{q_i} - r\right) \left\| f^{(1)} \right\|_{\left[r,\frac{p_i}{q_i}\right],1} \Big] \end{cases}$$

$$\leq \frac{1}{R-r} \times \begin{cases} \frac{\left\| f^{(1)} \right\|_{[r,R],\infty}}{2} \sum_{i=1}^m q_i \left[\left(R - \frac{p_i}{q_i}\right)^2 + \left(\frac{p_i}{q_i} - r\right)^2 \right] \\ \\ \frac{\left\| f^{(1)} \right\|_{[r,R],\beta}}{(\alpha+1)^{\frac{1}{\alpha}}} \sum_{i=1}^m q_i \left[\left(\frac{p_i}{q_i} - r\right)^{1+\frac{1}{\alpha}} + \left(R - \frac{p_i}{q_i}\right)^{1+\frac{1}{\alpha}} \right] \\ \text{if } \alpha > 1,\ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \\ (R-r) \left\| f^{(1)} \right\|_{[r,R],1} \end{cases}$$

$$\leq \begin{cases} \frac{\left\| f^{(1)} \right\|_{[r,R],\infty}(R-r)}{2} \\ \\ \frac{\left\| f^{(1)} \right\|_{[r,R],\beta}(R-r)^{\frac{1}{\alpha}}}{(\alpha+1)^{\frac{1}{\alpha}}} \quad \text{if } \alpha > 1,\ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \\ \left\| f^{(1)} \right\|_{[r,R],1} \end{cases}$$

which improves some results from [**68**].

CHAPTER 18

Two Functions Associated to f-Divergences

1. Introduction

In [**102**], Lin and Wong (see also [**101**]) introduced the following divergence measure

$$D_{LW}(p,q) := \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{\frac{1}{2}p_i + \frac{1}{2}q_i}\right), \quad p,q \in \mathbb{P}^n. \tag{1.1}$$

This can be represented as follows, using the Kullback-Leibler divergence:

$$D_{LW}(p,q) = D_{KL}\left(p, \frac{1}{2}p + \frac{1}{2}q\right).$$

Lin and Wong have established the following inequalities

$$D_{LW}(p,q) \le \frac{1}{2} D_{KL}(p,q); \tag{1.2}$$

$$D_{LW}(p,q) + D_{LW}(q,p) \le D_v(p,q) \le 2; \tag{1.3}$$

$$D_{LW}(p,q) \le 1. \tag{1.4}$$

In [**121**], Shioya and Da-te improved (1.2)-(1.4) by showing that

$$D_{LW}(p,q) \le \frac{1}{2} D_v(p,q) \le 1.$$

In the same paper [**121**], the authors have introduced the generalised Lin-Wong $f-$divergence $D_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right)$ and the Hermite-Hadamard divergence

$$D_{HH}^{f}(p,q) := \sum_{i=1}^{n} p_i \frac{\int_1^{\frac{q_i}{p_i}} f(t)\,dt}{\frac{q_i}{p_i} - 1}, \quad p,q \in \mathbb{P}^n$$

and, by the use of the Hermite-Hadamard inequality for convex functions, proved the following basic inequality

$$D_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right) \le D_{HH}^{f}(p,q) \le \frac{1}{2} D_f(p,q), \tag{1.5}$$

provided that f is convex and normalised, i.e., $f(1) = 0$.

In the following we point out new inequalites for $HH-$divergence, which also improve the above result.

2. Some Inequalities

In the following, we assume everywhere that the mapping $f : (0,\infty) \to \mathbb{R}$ is convex and normalised.

The following result holds.

THEOREM 2.1 (Barnett et al., 2003 [**3**]). *Let* $p, q \in \mathbb{P}^n$. *Then we have the inequality*

$$
\begin{aligned}
&D_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right) \\
&\le \lambda D_f\left(p, p + \frac{\lambda}{2}(q-p)\right) + (1-\lambda) D_f\left(p, \frac{p+q}{2} + \frac{\lambda}{2}(q-p)\right) \\
&\le D_{HH}^f(p,q) \le \frac{1}{2}\left[D_f(p,(1-\lambda)p + \lambda q) + (1-\lambda) D_f(p,q)\right] \\
&\le \frac{1}{2} D_f(p,q),
\end{aligned} \tag{2.1}
$$

for all $\lambda \in [0,1]$.

PROOF. Firstly, let us prove the following refinement of the Hermite-Hadamard inequality

$$
\begin{aligned}
&f\left(\frac{a+b}{2}\right) \\
&\le \lambda f\left(a + \lambda \cdot \frac{b-a}{2}\right) + (1-\lambda) f\left(\frac{a+b}{2} + \lambda \cdot \frac{b-a}{2}\right) \\
&\le \frac{1}{b-a}\int_a^b f(u)\,du \le \frac{1}{2}\left[f((1-\lambda)a + \lambda b) + \lambda f(a) + (1-\lambda) f(b)\right] \\
&\le \frac{f(a)+f(b)}{2}
\end{aligned} \tag{2.2}
$$

for all $\lambda \in [0,1]$.

Applying the Hermite-Hadamard inequality on every subinterval $[a,(1-\lambda)a+\lambda b], [(1-\lambda)a+\lambda b, b]$, we have

$$
\begin{aligned}
&f\left(\frac{a+(1-\lambda)a+\lambda b}{2}\right) \times [(1-\lambda)a + \lambda b - a] \\
&\le \int_a^{(1-\lambda)a+\lambda b} f(u)\,du \\
&\le \frac{f((1-\lambda)a+\lambda b) + f(a)}{2} \times [(1-\lambda)a + \lambda b - a]
\end{aligned}
$$

and

$$
\begin{aligned}
&f\left(\frac{(1-\lambda)a+\lambda b + b}{2}\right) \times [b - (1-\lambda)a - \lambda b] \\
&\le \int_{(1-\lambda)a+\lambda b}^b f(u)\,du \\
&\le \frac{f(b) + f((1-\lambda)a+\lambda b)}{2} \times [b - (1-\lambda)a - \lambda b],
\end{aligned}
$$

which are clearly equivalent to

$$
\begin{aligned}
\lambda f\left(a + \lambda \cdot \frac{b-a}{2}\right) &\le \frac{1}{b-a}\int_a^{(1-\lambda)a+\lambda b} f(u)\,du \\
&\le \frac{\lambda f((1-\lambda)a+\lambda b) + \lambda f(a)}{2}
\end{aligned} \tag{2.3}
$$

and

$$(1-\lambda) f\left(\frac{a+b}{2}+\lambda \cdot \frac{b-a}{2}\right) \leq \frac{1}{b-a} \int_{(1-\lambda) a+\lambda b}^{b} f(u) d u \leq \frac{(1-\lambda) f(b)+(1-\lambda) f((1-\lambda) a+\lambda b)}{2}. \tag{2.4}$$

Summing (2.3) and (2.4), we obtain the second and the first inequality in (2.2).

By the convexity property, we obtain

$$\begin{aligned} & \lambda f\left(a+\lambda \cdot \frac{b-a}{2}\right)+(1-\lambda) f\left(\frac{a+b}{2}+\lambda \cdot \frac{b-a}{2}\right) \\ & \geq f\left[\lambda\left(a+\lambda \cdot \frac{b-a}{2}\right)+(1-\lambda)\left(\frac{a+b}{2}+\lambda \cdot \frac{b-a}{2}\right)\right] \\ & =f\left(\frac{a+b}{2}\right) \end{aligned}$$

and the first inequality in (2.1) is proved.

The last inequality is obvious by the convexity property of f.

Now, if we choose $a=1$ and $b=\frac{q_i}{p_i}$, $x \in \chi$, in (2.2) and multiply by $p_i \geq 0$, $x \in \chi$, we get

$$\begin{aligned} & p_i f\left(\frac{p_i+q_i}{2 p_i}\right) \\ & \leq \lambda p_i f\left(\frac{p_i+\lambda\left(q_i-p_i\right)}{2 p_i}\right) \\ & +(1-\lambda) p_i f\left(\frac{p_i+q_i}{2 p_i}+\frac{\lambda\left(q_i-p_i\right)}{2 p_i}\right) \\ & \leq \frac{p^2(x)}{q_i-p_i} \int_1^{\frac{q_i}{p_i}} f(u) d u \\ & \leq \frac{1}{2}\left[f\left(\frac{(1-\lambda) p_i+\lambda q_i}{p_i}\right) p_i+\lambda p_i f(1)+(1-\lambda) p_i f\left(\frac{q_i}{p_i}\right)\right] \\ & \leq \frac{p_i f(1)+p_i f\left(\frac{q_i}{p_i}\right)}{2}. \end{aligned}$$

Integrating on χ and taking into account the definition of Csiszár $f-$divergence and the Hermite-Hadamard divergence, we obtain (2.1). ∎

REMARK 2.1. If $\lambda=0$ or $\lambda=1$, then by (2.1), we obtain the inequality (1.5).

COROLLARY 2.2 (Barnett et al., 2003 [**3**]). *Let $p, q \in \mathbb{P}^n$. Then we have the inequality*

$$\begin{aligned} D_f\left(p, \frac{p+q}{2}\right) & \leq \frac{1}{2}\left[D_f\left(p, \frac{3 p+q}{4}\right)+D_f\left(p, \frac{p+3 q}{4}\right)\right] \\ & \leq D_{H H}^f(p, q) \leq \frac{1}{2}\left[D_f\left(p, \frac{p+q}{2}\right)+\frac{1}{2} D_f(p, q)\right] \\ & \leq \frac{1}{2} D_f(p, q). \end{aligned} \tag{2.5}$$

REMARK 2.2. If we substitute λ by $(1-\lambda)$ in (2.1), we can get

$$\begin{aligned} D_f\left(p, \frac{p+q}{2}\right) & \\ &\leq (1-\lambda) D_f\left(p, \frac{p+q}{2} + \lambda(p-q)\right) + \lambda D_f\left(p, q + \lambda\frac{p-q}{2}\right) \\ &\leq D_{HH}^f(p,q) \leq \frac{1}{2}\left[D_f(p, \lambda p + (1-\lambda) q) + \lambda D_f(p,q)\right] \\ &\leq \frac{1}{2} D_f(p,q). \end{aligned} \tag{2.6}$$

Now, if we add (2.1) and (2.6) and divide by 2, we can state the following corollary.

COROLLARY 2.3 (Barnett et al., 2003 [**3**]). *Let* $p, q \in \mathbb{P}^n$. *Then we have the inequality*

$$\begin{aligned} D_f\left(p, \frac{p+q}{2}\right) & \\ &\leq \lambda\left[D_f\left(p, p + \frac{\lambda}{2}(q-p)\right) + D_f\left(p, q + \frac{\lambda}{2}(p-q)\right)\right] \\ &+ (1-\lambda)\left[D_f\left(p, \frac{p+q}{2} + \frac{\lambda}{2}(q-p)\right) + D_f\left(p, \frac{p+q}{2} + \frac{1}{2}(p-q)\right)\right] \\ &\leq D_{HH}^f(p,q) \\ &\leq \frac{1}{4}\left[D_f(p, (1-\lambda) p + \lambda q) + D_f(p, \lambda p + (1-\lambda) q) + D_f(p,q)\right] \\ &\leq \frac{1}{2} D_f(p,q), \end{aligned} \tag{2.7}$$

for all $\lambda \in [0,1]$.

We also define the divergence.

$$H_f(p,q;t) := \sum_{i=1}^{n} p_i f\left[\frac{tq_i + (1-t) p_i}{p_i}\right] = D_f(p, tq + (1-t) p). \tag{2.8}$$

We can state the following theorem.

THEOREM 2.4 (Barnett et al., 2003 [**3**]). *Let* $p, q \in \mathbb{P}^n$. *Then*

(i) $H_f(p,q;\cdot)$ *is convex on* $[0,1]$;

(ii) *We have the bounds*

$$\inf_{t\in[0,1]} H_f(p,q;t) = H_f(p,q;0) = 0, \tag{2.9}$$

$$\sup_{t\in[0,1]} H_f(p,q;t) = H_f(p,q;1) = D_f(p,q), \tag{2.10}$$

and the inequality

$$H_f(p,q;t) \leq tD_f(p,q) \quad \textit{for all } t \in [0,1]. \tag{2.11}$$

(iii) *The mapping* $H_f(p,q;\cdot)$ *is monotonic nondecreasing on* $[0,1]$.

PROOF. We have:

(i) Let t_1, $t_2 \in [0,1]$ and $\alpha, \beta \in [0,1]$ with $\alpha + \beta = 1$. Then

$$\begin{aligned}
&H_f(p,q;\alpha t_1 + \beta t_2) \\
&= \sum_{i=1}^{n} p_i f\left[\frac{(\alpha t_1 + \beta t_2) q_i + (1 - \alpha t_1 - \beta t_2) p_i}{q_i}\right] \\
&= \sum_{i=1}^{n} p_i f\left[\alpha \cdot \frac{[t_1 q_i + (1-t_1) p_i]}{q_i} + \beta \cdot \frac{[t_2 q_i + (1-t_2) p_i]}{q_i}\right] \\
&\leq \alpha \cdot \sum_{i=1}^{n} p_i f\left[\frac{[t_1 q_i + (1-t_1) p_i]}{q_i}\right] \\
&+ \beta \cdot \sum_{i=1}^{n} \frac{[t_2 q_i + (1-t_2) p_i]}{q_i} \\
&= \alpha H_f(p,q,t_1) + \beta H_f(p,q,t_2)
\end{aligned}$$

and the convexity is proved.

(ii) Using Jensen's inequality, we have:

$$\begin{aligned}
H_f(p,q,t) &\geq f\left[\sum_{i=1}^{n} p_i \left[\frac{t q_i + (1-t) p_i}{q_i}\right]\right] \\
&= f\left[t \sum_{i=1}^{n} q_i + (1-t) \sum_{i=1}^{n} p_i\right] \\
&= f(1) = 0 = H_f(p,q,0).
\end{aligned}$$

Also, by the convexity of f, we have

$$\begin{aligned}
H_f(p,q,t) &\leq \sum_{i=1}^{n} p_i \left[t f\left(\frac{q_i}{p_i}\right) + (1-t) f(1)\right] \\
&\leq t \sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right) + (1-t) f(1) \sum_{i=1}^{n} p_i \\
&= t D_f(p,q),
\end{aligned}$$

and the statement (ii) is proved.

(iii) Let t_1, $t_2 \in [0,1]$ with $t_2 > t_1$. As $H_f(p,q;\cdot)$ is convex, then

$$\frac{H_f(p,q,t_2) - H_f(p,q,t_1)}{t_2 - t_1} \geq \frac{H_f(p,q,t_1) - H_f(p,q,0)}{t_1 - 0}$$

and as

$$H_f(p,q,t_1) \geq H_f(p,q,0) = 0,$$

we deduce that $H_f(p,q,t_1) \geq H_f(p,q,t_2)$, which proves that monotonicity of $H_f(p,q,\cdot)$. ∎

REMARK 2.3. If we write (2.11) for $1-t$, we obtain

(2.12) $$H_f(p,q,1-t) \leq (1-t) D_f(p,q), \quad t \in [0,1].$$

Then, adding (2.11) and (2.12), we get

(2.13) $$H_f(p,q,t) + H_f(p,q,1-t) \leq D_f(p,q)$$

for all $t \in [0,1]$.

REMARK 2.4. For $t \in \left[\frac{1}{2}, 1\right]$, we have the inequality

$$D_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right) \leq D_f\left(p, tq + (1-\lambda)p\right) \leq tD_f(p,q), \tag{2.14}$$

which is similar with (1.3).

We can also define the divergence

$$F_f(p,q;t) := \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j f\left[t \cdot \frac{q_i}{p_i} + (1-t)\cdot \frac{q_j}{p_j}\right], \tag{2.15}$$

where $p, q \in \mathbb{P}^n$ and $t \in [0,1]$.

The properties of this mapping are embodied in the following theorem.

THEOREM 2.5 (Barnett et al., 2003 [**3**]). *Let* $p, q \in \mathbb{P}^n$. *Then:*

(i) $F_f(p,q;\cdot)$ *is symmetrical about* $\frac{1}{2}$*, i.e.,*

$$F_f(p,q;t) = F_f(p,q;1-t) \quad \textit{for all } t \in [0,1]. \tag{2.16}$$

(ii) F *is convex on* $[0,1]$*;*
(iii) *We have the bounds:*

$$\sup_{t\in[0,1]} F_f(p,q;t) = F_f(p,q;0) = F_f(p,q;1) = D_f(p,q), \tag{2.17}$$

$$\begin{aligned}\inf_{t\in[0,1]} F_f(p,q;t) &= F_f\left(p,q;\frac{1}{2}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j f\left(\frac{q_i p_j + p_i q_j}{2p_i q_j}\right) \geq 0\end{aligned} \tag{2.18}$$

(iv) $F_f(p,q;\cdot)$ *is nondecreasing on* $\left[0, \frac{1}{2}\right]$ *and nonincreasing on* $\left[\frac{1}{2}, 1\right]$*;*
(v) *We have the inequality:*

$$F_f(p,q;t) \geq \max\left\{H_f(p,q;t); H_f(p,q;1-t)\right\} \quad \textit{for all } \ t \in [0,1]. \tag{2.19}$$

PROOF. We have:
(i) It is obvious.
(ii) Follows by the convexity of f in a similar way to that in the proof of Theorem 2.4.
(iii) For all $x, y \in \chi$ we have:

$$f\left[t \cdot \frac{q_i}{p_i} + (1-t)\cdot \frac{q_j}{p_j}\right] \leq t \cdot f\left(\frac{q_i}{p_i}\right) + (1-t)\cdot f\left(\frac{q_j}{p_j}\right)$$

for any $t \in [0,1]$.
Multiplying by $p_i p_j \geq 0$ and integrating on χ^2, we may write

$$\begin{aligned} F_f(p,q;t) &\leq \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j \left[t \cdot f\left(\frac{q_i}{p_i}\right) + (1-t)\cdot f\left(\frac{q_j}{p_j}\right)\right] \\ &= t\sum_{j=1}^{n} p_j \sum_{i=1}^{n} p_i f\left(\frac{q_i}{p_i}\right) \\ &+ (1-t)\sum_{i=1}^{n}\sum_{j=1}^{n} p_j f\left(\frac{q_j}{p_j}\right) \\ &= t\cdot D_f(p,q) + (1-t)\cdot D_f(p,q) = D_f(p,q) \\ &= F_f(p,q;0) = F_f(p,q;1) \end{aligned}$$

and the bound (2.17) is proved.
Since f is convex, then for all $t \in [0,1]$ and $x, y \in \chi$, we have

$$\begin{aligned} &\frac{1}{2}\left\{ f\left[t\cdot\frac{q_i}{p_i} + (1-t)\cdot\frac{q_j}{p_j}\right] + f\left[(1-t)\cdot\frac{q_i}{p_i} + t\cdot\frac{q_j}{p_j}\right]\right\} \\ &\geq f\left[\frac{1}{2}\left(\frac{q_i}{p_i}+\frac{q_j}{p_j}\right)\right]. \end{aligned}$$

Multiplying by $p_i p_j \geq 0$ and integrating on χ^2, we have

$$\frac{1}{2}\left[F_f(p,q;t) + F_f(p,q;1-t)\right] \geq \sum_{i=1}^{n}\sum_{j=1}^{n} p_i p_j f\left[\frac{1}{2}\left(\frac{q_i}{p_i}+\frac{q_j}{p_j}\right)\right]$$

and the first part of (2.18) is proved.
Using Jensen's integral inequality, we may write:

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{n} f\left[\frac{1}{2}\left(\frac{q_i p_j + p_i q_j}{p_i q_j}\right)\right] p_i p_j \\ &\geq f\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{2}\left(\frac{q_i p_j + p_i q_j}{p_i q_j}\right) p_i p_j\right] \\ &= f\left[\frac{1}{2}\left[\sum_{i=1}^{n} p_i \sum_{j=1}^{n} p_j + \sum_{i=1}^{n} q_i \sum_{j=1}^{n} q_j\right]\right] = f(1) = 0 \end{aligned}$$

and the second part of (2.18) is proved.

(iv) The mapping $F_f(p,q;\cdot)$ being convex on $[0,1]$, we may write for $1 \geq t_2 > t_1 \geq \frac{1}{2}$ that

$$\frac{F_f(p,q;t_2) - F_f(p,q;t_1)}{t_2 - t_1} \geq \frac{F_f(p,q;t_1) - F_f\left(p,q;\frac{1}{2}\right)}{t_1 - \frac{1}{2}}$$

and as

$$F_f(p,q;t_1) \geq F_f\left(p,q;\frac{1}{2}\right), \quad t_1 \geq \frac{1}{2},$$

we deduce that $F_f(p,q;t_2) \geq F_f(p,q;t_1)$, i.e., the mapping $F_f(p,q;\cdot)$ is monotonically nondecreasing on $\left[0,\frac{1}{2}\right]$.
Similarly, we can prove that $F_f(p,q;\cdot)$ is monotonically nonincreasing on $\left[0,\frac{1}{2}\right]$, and the statement (iv) is proved.

(v) Using Jensen's integral inequality, we have

$$\sum_{j=1}^{n} p_j f\left[t \cdot \frac{q_i}{p_i} + (1-t) \cdot \frac{q_j}{p_j}\right]$$
$$\geq f\left[\sum_{j=1}^{n} p_j \left[t \cdot \frac{q_i}{p_i} + (1-t) \cdot \frac{q_j}{p_j}\right]\right]$$
$$= f\left[t \cdot \frac{q_i}{p_i} \sum_{j=1}^{n} p_j + (1-t) \cdot \sum_{j=1}^{n} q_j\right] = f\left[t \cdot \frac{q_i}{p_i} + (1-t)\right].$$

Multiplying by $p_i \geq 0$ and integrating on χ, we have

$$F_f(p,q;t) \geq \sum_{i=1}^{n} p_i f\left[t \cdot \frac{q_i}{p_i} + (1-t)\right] = H_f(p,q;t),$$

for all $t \in [0,1]$.
Now, as

$$F_f(p,q;1-t) \geq H_f(p,q;1-t)$$

and $F_f(p,q;t) = F_f(p,q;1-t)$ for all $t \in [0,1]$, the inequality (2.19) is completely proved. ∎

3. Preliminary Results

In [3], the authors introduced the following divergence measure

$$H_f(p,q;t) := \sum_{i=1}^{m} p_i f\left[\frac{tq_i + (1-t)p_i}{p_i}\right], \tag{3.1}$$

where $p, q \in Q$ and $t \in [0,1]$.

It is obvious that this measure can be represented in terms of $f-$divergence, namely, we have the representation,

$$H_f(p,q;t) = I_f(p, tq + (1-t)p) \tag{3.2}$$

for all $p, q \in Q$ and $t \in [0,1]$.

The following properties of $H_f(\cdot,\cdot;\cdot)$ hold (see [3]).

THEOREM 3.1. *Assume that the mapping* $f : [0,\infty) \to \mathbb{R}$ *is convex and* $p, q \in Q$*, then,*

(i) $H_f(p,q;\cdot)$ *is convex on* $[0,1]$*;*

(ii)

$$H_f(p,q;t) \leq I_f(p,q) \quad \textit{for all } t \in [0,1] \tag{3.3}$$

with the bounds

$$\inf_{t\in[0,1]} H_f(p,q;t) = H_f(p,q;0) = 0 \tag{3.4}$$

and

$$\sup_{t\in[0,1]} H_f(p,q;t) = H_f(p,q;1) = I_f(p,q); \tag{3.5}$$

(iii) *The mapping* $H_f(p,q;\cdot)$ *is monotonic nondecreasing on* $[0,1]$.

In the same paper [**3**], the authors introduced the following divergence,

$$F_f(p,q;t)=\sum_{i=1}^{m}\sum_{j=1}^{m}p_ip_jf\left[t\cdot\frac{q_i}{p_i}+(1-t)\cdot\frac{q_j}{p_j}\right], \tag{3.6}$$

where $p,q\in\Omega$ and $t\in[0,1]$.

The properties of this mapping are embodied in the following theorem [**3**].

THEOREM 3.2 (Barnett et al., 2003 [**3**]). *Under the assumptions of Theorem 3.1, we have,*

(i) $F_f(p,q;\cdot)$ *is symmetrical about* $\frac{1}{2}$*, i.e.,*

$$F_f(p,q;t)=F_f(p,q;1-t)\ \ \textit{for all } t\in[0,1]\,; \tag{3.7}$$

(ii) $F_f(p,q;\cdot)$ *is convex on* $[0,1]$*;*

(iii) *We have the bounds*

$$\sup_{t\in[0,1]}F_f(p,q;t)=F_f(p,q;0)=F_f(p,q;1)=I_f(p,q)\,; \tag{3.8}$$

$$\begin{aligned}\inf_{t\in[0,1]}F_f(p,q;t)&=F_f\left(p,q;\frac{1}{2}\right)\\&=\sum_{i=1}^{m}\sum_{j=1}^{m}p_ip_jf\left[\frac{q_ip_j+p_iq_j}{2p_ip_j}\right]\geq 0;\end{aligned} \tag{3.9}$$

(iv) $F_f(p,q;\cdot)$ *is nondecreasing on* $\left[0,\frac{1}{2}\right]$ *and nonincreasing on* $\left[\frac{1}{2},1\right]$*;*

(v) *and*

$$F_f(p,q;t)\geq\max\left\{H_f(p,q;t),H_f(p,q;1-t)\right\} \tag{3.10}$$

for all $t\in[0,1]$.

In this section we point out some estimates for the divergence measures $F_f(\cdot,\cdot;\cdot)$ and $H_f(\cdot,\cdot;\cdot)$.

4. SOME ESTIMATES FOR n-TIME DIFFERENTIABLE MAPPINGS

We use the following lemma (see also [**33**]).

LEMMA 4.1. *Let* $f:I\in\mathbb{R}\to\mathbb{R}$ *(*I *interval of* $\mathbb{R}$*) be such that* $f^{(n)}$ *is absolutely continuous on* I*, then for all* $x,a\in\mathring{I}$ *(*$\mathring{I}$ *is the interior of* I*) we have the inequality,*

$$\left|f(x)-f(a)-\sum_{k=1}^{n}\frac{(x-a)^k}{k!}f^{(k)}(a)\right| \leq\begin{cases}\frac{1}{(n+1)!}\left\|f^{(n+1)}\right\|_\infty|x-a|^{n+1} & \textit{if}\quad f^{(n+1)}\in L_\infty(I)\,;\\ \frac{1}{n!(n\beta+1)^{\frac{1}{\beta}}}\left\|f^{(n+1)}\right\|_\alpha|x-a|^{n+\frac{1}{\beta}} & \textit{if}\quad f^{(n+1)}\in L_\alpha(I)\,,\\ & \qquad\alpha>1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1;\\ \frac{1}{n!}\left\|f^{(n+1)}\right\|_1|x-a|^n\,, & \end{cases} \tag{4.1}$$

where $\|\cdot\|_\alpha$ $(\alpha\in[1,\infty])$ *are the usual Lebesgue norms on* I*, i.e.,*

$$\|g\|_\alpha:=\left(\int_I|g(x)|^\alpha dx\right)^{\frac{1}{\alpha}},\quad\alpha\geq 1$$

$$\|g\|_\infty:=ess\sup_{x\in I}|g(x)|\,.$$

The following corollary will be useful in what follows.

COROLLARY 4.2. *Assume that f is as above and $a, b \in \mathring{I}$, then for all $\lambda \in [0,1]$, we have the inequality:*

$$\left| f(\lambda b + (1-\lambda) a) - f(a) - \sum_{k=1}^{n} \frac{\lambda^k (b-a)^k}{k!} f^{(k)}(a) \right| \tag{4.2}$$
$$\leq \begin{cases} \frac{\lambda^{n+1}|b-a|^{n+1}}{(n+1)!} \left\| f^{(n+1)} \right\|_{\infty} & \text{if } f^{(n+1)} \in L_{\infty}(I); \\ \frac{\lambda^{n+\frac{1}{\beta}}|b-a|^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_{\infty} & \text{if } f^{(n+1)} \in L_{\alpha}(I), \\ & \alpha > 1,\ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \frac{\lambda^{n}|b-a|^{n}}{n!} \left\| f^{(n+1)} \right\|_{1}. \end{cases}$$

We can now point out the following estimation result for the mapping $H_f(p, q; \cdot)$ [**6**].

THEOREM 4.3 (Barnett et al., 2003 [**6**]). *Assume that the mapping $f : [0, \infty) \to \mathbb{R}$ is such that $f^{(n)}$ is absolutely continuous on $[r, R]$, where $0 \leq r \leq 1 \leq R < \infty$. If p, q are probability distributions and*

$$r \leq \frac{q_i}{p_i} \leq R \text{ for } i \in \{1, ..., m\}, \tag{4.3}$$

then we have,

$$\left| H_f(p, q; t) - f(1) - \sum_{k=1}^{n} \frac{t^k f^{(k)}(1)}{k!} D_{\chi^k}(p, q) \right| \tag{4.4}$$
$$\leq \begin{cases} \frac{t^{n+1}\left\| f^{(n+1)} \right\|_{\infty}}{(n+1)!} D_{|\chi|^{n+1}}(p, q) & \text{if } f^{(n+1)} \in L_{\infty}[r, R]; \\ \frac{t^{n+\frac{1}{\beta}}\left\| f^{(n+1)} \right\|_{\alpha}}{n!(n\beta+1)^{\frac{1}{\beta}}} D_{|\chi|^{n+\frac{1}{\beta}}}(p, q) & \text{if } f^{(n+1)} \in L_{\alpha}[r, R], \\ & \alpha > 1,\ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \frac{t^{n}\left\| f^{(n+1)} \right\|_{1}}{n!} D_{|\chi|^{n}}(p, q), \end{cases}$$
$$\leq \begin{cases} \frac{t^{n+1}(R-r)^{n+1}}{(n+1)!} \left\| f^{(n+1)} \right\|_{\infty} \\ \frac{t^{n+\frac{1}{\beta}}(R-r)^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_{\alpha} \\ \frac{t^{n}(R-r)^{n}}{n!} \left\| f^{(n+1)} \right\|_{1}, \end{cases}$$

where

$$D_{\chi^k}(p, q) := \sum_{i=1}^{m} \frac{(q_i - p_i)^k}{p_i^{k-1}}, \quad k = 1, ...$$

and

$$D_{|\chi|^r}(p, q) := \sum_{i=1}^{m} \frac{|q_i - p_i|^r}{p_i^{r-1}}, \quad r \geq 0$$

where the Lebesgue $\alpha-$norms are taken on $[r, R]$.

PROOF. Apply inequality (4.1) for $\lambda = t \in [0,1]$, $b = \frac{q_i}{p_i}$, $i \in \{1,...,m\}$ and $a = 1$, to get,

$$\left| f\left(t \cdot \frac{q_i}{p_i} + (1-t)\right) - f(1) - \sum_{k=1}^{n} \frac{t^k \left(\frac{q_i}{p_i} - 1\right)^k}{k!} f^{(k)}(1) \right| \tag{4.5}$$
$$\leq \begin{cases} \frac{t^{n+1}\left|\frac{q_i}{p_i}-1\right|^{n+1}}{(n+1)!} \left\| f^{(n+1)} \right\|_{\infty}; \\ \frac{t^{n+\frac{1}{\beta}}\left|\frac{q_i}{p_i}-1\right|^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_{\alpha}; \\ \frac{t^{n}\left|\frac{q_i}{p_i}-1\right|^{n}}{n!} \left\| f^{(n+1)} \right\|_{1}; \end{cases} \leq \begin{cases} \frac{t^{n+1}(R-r)^{n+1}}{(n+1)!} \left\| f^{(n+1)} \right\|_{\infty}; \\ \frac{t^{n+\frac{1}{\beta}}(R-r)^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_{\alpha}; \\ \frac{t^{n}(R-r)^{n}}{n!} \left\| f^{(n+1)} \right\|_{1}, \end{cases}$$

for $i \in \{1,...,m\}$.

If we multiply (4.5) by $p_i \geq 0$, sum over i and use the properties of the sum, we get,

$$\left| \sum_{i=1}^{m} p_i f\left(\frac{tq_i + (1-t)p_i}{p_i}\right) - f(1) - \sum_{k=1}^{n} \frac{t^k f^{(k)}(1)}{k!} \sum_{i=1}^{m} \frac{(q_i - p_i)^k}{p_i^{k-1}} \right|$$

$$\leq \begin{cases} \frac{t^{n+1}\left\| f^{(n+1)} \right\|_{\infty}}{(n+1)!} \sum_{i=1}^{m} \frac{|q_i - p_i|^{n+1}}{p_i^{n}}; \\ \frac{t^{n+\frac{1}{\beta}}\left\| f^{(n+1)} \right\|_{\alpha}}{n!(n\beta+1)^{\frac{1}{\beta}}} \sum_{i=1}^{m} \frac{|q_i - p_i|^{n+\frac{1}{\beta}}}{p_i^{n+\frac{1}{\beta}-1}}; \\ \frac{t^{n}\left\| f^{(n+1)} \right\|_{1}}{n!} \sum_{i=1}^{m} \frac{|q_i - p_i|^{n}}{p_i^{n-1}}, \end{cases}$$
$$\leq \begin{cases} \frac{t^{n+1}(R-r)^{n+1}}{(n+1)!} \left\| f^{(n+1)} \right\|_{\infty} \\ \frac{t^{n+\frac{1}{\beta}}(R-r)^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_{\alpha} \\ \frac{t^{n}(R-r)^{n}}{n!} \left\| f^{(n+1)} \right\|_{1}, \end{cases}$$

and the theorem is proved. ∎

REMARK 4.1. If $n = 0$, then, basically, for an absolutely continuous mapping $f : [r, R] \subset [0, \infty) \to \mathbb{R}$, we have:

$$|I_f(p, tq + (1-t)p) - f(1)| \leq \begin{cases} t\|f'\|_{\infty} D_v(p,q) \\ t^{\frac{1}{\beta}} \|f'\|_{\alpha} D_{|\alpha|^{\frac{1}{\beta}}}(p,q) \\ \|f'\|_1 \end{cases} \leq \begin{cases} t\|f'\|_{\infty}(R-r) \\ t^{\frac{1}{\beta}}(R-r)^{\frac{1}{\beta}} \cdot \|f'\|_{\alpha} \\ \|f'\|_1 \end{cases} \tag{4.6}$$

for all $t \in [0,1]$, where $D_v(p,q) = \sum_{i=1}^{m} |p_i - q_i|$.

If $n = 1$, and taking into account that $D_{\chi}(p,q) = 0$, then by (4.4) we get, for the mappings whose derivatives f' are absolutely continuous,

$$|I_f(p, tq + (1-t)p) - f(1)| \tag{4.7}$$
$$\leq \begin{cases} \frac{t^2\|f'\|_\infty}{2} D_{\chi^2}(p,q) & \text{if} \quad f' \in L_\infty[r,R] \\ \frac{t^{\frac{\beta+1}{\beta}}\|f'\|_\alpha}{(\beta+1)^{\frac{1}{\beta}}} D_{|\chi|^{\frac{\beta+1}{\beta}}}(p,q) & \text{if} \quad f' \in L_\alpha[r,R] \\ t\|f'\|_1 D_v(p,q) \end{cases}$$

for all $t \in [0,1]$.

Of course, if we assume that f is convex and normalised, then the left hand side of both (4.6) and (4.7) will become

$$0 \leq I_f(p, tq + (1-t)p)$$

and the inequalities (4.6) and (4.7) will provide some upper bounds for the mapping $H_f(p,q;t)$, $t \in [0,1]$.

REMARK 4.2. If we assume that f'' is absolutely continuous, then from (4.4) we obtain,

$$\left|H_f(p,q;t) - f(1) - \frac{t^2}{2} f''(1) D_{\chi^2}(p,q)\right| \tag{4.8}$$
$$\leq \begin{cases} \frac{t^3\|f'''\|_\infty}{6} D_{|\chi|^3}(p,q) & \text{if} \quad f''' \in L_\infty[r,R]; \\ \frac{t^{\frac{2\beta+1}{\beta}}\|f'''\|_\alpha}{2(2\beta+1)^{\frac{1}{\beta}}} D_{|\chi|^{2+\frac{1}{\beta}}}(p,q) & \text{if} \quad f''' \in L_\alpha[r,R], \\ & \alpha > 1, \ \frac{1}{\alpha} + \frac{1}{\beta} = 1; \\ \frac{t^2\|f'''\|_1}{2} D_{|\chi|^2}(p,q), \end{cases}$$

which provides an approximation of $H_f(p,q;t)$ by a quadratic in t whose coefficient is dependent on the χ^2 -distance of p and q.

We also note that Theorem 4.3 contains, as a particular case (for $t = 1$), an approximation of the $f-$divergence dontained in the following corollary.

COROLLARY 4.4. *With the assumptions of Theorem 4.3, we have,*

$$\left|I_f(p,q) - f(1) - \sum_{k=1}^{n} \frac{f^{(k)}(1)}{k!} D_{\chi^k}(p,q)\right| \tag{4.9}$$
$$\leq \begin{cases} \frac{\left\|f^{(n+1)}\right\|_\infty}{(n+1)!} D_{|\chi|^{n+1}}(p,q) \\ \frac{\left\|f^{(n+1)}\right\|_\alpha}{n!(n\beta+1)^{\frac{1}{\beta}}} D_{|\chi|^{n+\frac{1}{\beta}}}(p,q) \\ \frac{\left\|f^{(n+1)}\right\|_1}{n!} D_{|\chi|^n}(p,q) \end{cases} \leq \begin{cases} \frac{\left\|f^{(n+1)}\right\|_\infty}{(n+1)!} (R-r)^{n+1} \\ \frac{\left\|f^{(n+1)}\right\|_\alpha}{n!(n\beta+1)^{\frac{1}{\beta}}} (R-r)^{n+\frac{1}{\beta}} \\ \frac{\left\|f^{(n+1)}\right\|_1}{n!} (R-r)^n. \end{cases}$$

We also know that for $t = \frac{1}{2}$, we obtain the generalised Lin-Wong $f-$divergence

$$LW_f(p,q) := I_f\left(p, \frac{1}{2}p + \frac{1}{2}q\right)$$

and so, from (4.4), we may state the following estimation for the Lin-Wong $f-$divergence.

COROLLARY 4.5. *With the assumptions of Theorem 4.3, we have,*

$$\left|LW_f(p,q) - f(1) - \sum_{k=1}^{n} \frac{t^k f^{(k)}(1)}{2^k k!} D_{\chi^k}(p,q)\right| \tag{4.10}$$

$$\leq \begin{cases} \frac{\left\|f^{(n+1)}\right\|_\infty}{2^{n+1}(n+1)!} D_{|\chi|^{n+1}}(p,q) & \text{if } f^{(n+1)} \in L_\infty[r,R]; \\ \frac{\left\|f^{(n+1)}\right\|_\alpha}{2^{n+\frac{1}{\beta}} n!(n\beta+1)^{\frac{1}{\beta}}} D_{|\chi|^{n+\frac{1}{\beta}}}(p,q) & \text{if } f^{(n+1)} \in L_\alpha[r,R], \\ & \alpha > 1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{\left\|f^{(n+1)}\right\|_1}{2^n n!} D_{|\chi|^n}(p,q). \end{cases}$$

REMARK 4.3. Similar particular cases for $n = 0$, $n = 1$ and $n = 2$ may be stated but we omit the details.

The following theorem also holds [**6**].

THEOREM 4.6 (Barnett et al., 2003 [**6**]). *Assume that the mapping $f : [0,\infty) \to \mathbb{R}$ is such that $f^{(n)}$ is absolutely continuous on $[r,R]$, where $0 \leq r \leq 1 \leq R < \infty$. If p,q satisfy the condition*

$$r \leq \frac{q_i}{p_i} \leq R \text{ for } i \in \{1,...,m\}, \tag{4.11}$$

then we have the inequality,

$$\left|F_f(p,q;t) - I_f(p,q) - \sum_{k=1}^{n} \frac{t^k}{k!} D^{(*)}_{f^{(k)}}(p,q)\right| \tag{4.12}$$

$$\leq \begin{cases} \frac{t^{n+1}}{(n+1)!} D^{(*)}_{n+1}(p,q) \left\|f^{(n+1)}\right\|_\infty & \text{if } f^{(n+1)} \in L_\infty[r,R]; \\ \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} D^{(*)}_{n+\frac{1}{\beta}}(p,q) \left\|f^{(n+1)}\right\|_\alpha & \text{if } f^{(n+1)} \in L_\alpha[r,R], \\ & \alpha > 1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{t^n}{n!} D^{(*)}_n(p,q) \left\|f^{(n+1)}\right\|_1 \end{cases}$$

$$\leq \begin{cases} \frac{t^{n+1}}{(n+1)!} (R-r)^{n+1} \left\|f^{(n+1)}\right\|_\infty & \text{if } f^{(n+1)} \in L_\infty[r,R]; \\ \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} (R-r)^{n+\frac{1}{\beta}} \left\|f^{(n+1)}\right\|_\alpha & \text{if } f^{(n+1)} \in L_\alpha[r,R], \\ & \alpha > 1,\ \frac{1}{\alpha}+\frac{1}{\beta}=1; \\ \frac{t^n}{n!} (R-r)^n \left\|f^{(n+1)}\right\|_1, \end{cases}$$

where

$$D^{(*)}_{f^{(k)}}(p,q) = \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\left(\det\begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right)^k}{[p_i]^{k-1}[p_j]^{k-1}} f^{(k)}\left(\frac{q_j}{p_j}\right), \quad k = 1, ...$$

$$D^{(*)}_s(p,q) = \sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\left|\det\begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right|^s}{p_i^{s-1} p_j^{s-1}}, \quad s > 0$$

and the $\alpha-$norms are taken on $[r,R]$.

PROOF. We choose in Corollary 4.2, $b = \frac{q_i}{p_i}$, $a = \frac{q_j}{p_j}$, to obtain,

$$\left| f\left(t \cdot \frac{q_i}{p_i} + (1-t) \cdot \frac{q_j}{p_j}\right) - f\left(\frac{q_j}{p_j}\right) - \sum_{k=1}^{n} \frac{t^k \left(\frac{q_i}{p_i} - \frac{q_j}{p_j}\right)^k}{k!} f^{(k)}\left(\frac{q_j}{p_j}\right) \right| \leq \begin{cases} \frac{t^{n+1}}{(n+1)!} \left|\frac{q_i}{p_i} - \frac{q_j}{p_j}\right|^{n+1} \left\|f^{(n+1)}\right\|_\infty \\ \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left|\frac{q_i}{p_i} - \frac{q_j}{p_j}\right|^{n+\frac{1}{\beta}} \left\|f^{(n+1)}\right\|_\alpha \\ \frac{t^n}{n!} \left|\frac{q_i}{p_i} - \frac{q_j}{p_j}\right|^{n} \left\|f^{(n+1)}\right\|_1 \end{cases}$$

for all $i, j \in \{1, ..., m\}$ and $t \in [0, 1]$, which is clearly equivalent to,

$$\left| f\left(\frac{tp_jq_i + (1-t)\,p_iq_j}{p_ip_j}\right) - f\left(\frac{q_j}{p_j}\right) - \sum_{k=1}^{n} \frac{t^k}{k!} \cdot \frac{\left(\det \begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right)^k}{p_i^k p_j^k} f^{(k)}\left(\frac{q_j}{p_j}\right) \right| \leq \begin{cases} \frac{t^{n+1}}{(n+1)!} \cdot \frac{\left|\det \begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right|^{n+1}}{p_i^{n+1} p_j^{n+1}} \left\|f^{(n+1)}\right\|_\infty \\ \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \cdot \frac{\left|\det \begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right|^{n+\frac{1}{\beta}}}{p_i^{n+\frac{1}{\beta}} p_j^{n+\frac{1}{\beta}}} \left\|f^{(n+1)}\right\|_\alpha \\ \frac{t^n}{n!} \cdot \frac{\left|\det \begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right|^{n}}{p_i^{n} p_j^{n}} \left\|f^{(n+1)}\right\|_1 \end{cases} \tag{4.13}$$

for all $i, j \in \{1, ..., m\}$ and $t \in [0, 1]$.

If we multiply (4.13) by $p_i p_j \geq 0$ for $i, j \in \{1, ..., m\}$, sum over i and j, we obtain,

$$\left| \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j f\left(\frac{t p_j q_i + (1-t) p_i q_j}{p_i p_j}\right) - \sum_{i=1}^{m} \sum_{j=1}^{m} p_i p_j f\left(\frac{q_j}{p_j}\right) - \sum_{k=1}^{n} \frac{t^k}{k!} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\left(\det \begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right)^k}{p_i^{k-1} p_j^{k-1}} f^{(k)}\left(\frac{q_j}{p_j}\right) \right|$$
$$\leq \begin{cases} \frac{t^{n+1}}{(n+1)!} \left\| f^{(n+1)} \right\|_\infty \cdot \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\left|\det \begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right|^{n+1}}{p_i^n p_j^n} \\ \frac{t^{n+\frac{1}{\beta}}}{n!(n\beta+1)^{\frac{1}{\beta}}} \left\| f^{(n+1)} \right\|_\alpha \cdot \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\left|\det \begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right|^{n+\frac{1}{\beta}}}{p_i^{n+\frac{1}{\beta}-1} p_j^{n+\frac{1}{\beta}-1}} \\ \frac{t^n}{n!} \left\| f^{(n+1)} \right\|_1 \cdot \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\left|\det \begin{bmatrix} p_j & q_j \\ p_i & q_i \end{bmatrix}\right|^{n}}{p_i^{n-1} p_j^{n-1}}, \end{cases}$$

which is clearly equivalent to the first inequality in (4.12).

The second inequality is obvious by the fact that,

$$\left| \frac{q_i}{p_i} - \frac{q_j}{p_j} \right| \leq R - r \text{ for all } i, j \in \{1, ..., m\}.$$

The theorem is thus completely proved. ∎

REFERENCES

[1] S. M. ALI and S. D. SILVEY, A general class of coefficients of divergence of one distribution from another, *J. Roy. Statist. Soc. Sec B*, **28** (1966), 131-142.

[2] D. ANDRICA and I. RAŞA, The Jensen inequality: refinements and applications, *Anal. Numer. Theor. Approx.*, **74** (1985), 105-108.

[3] N. S. BARNETT, P. CERONE and S. S. DRAGOMIR, Some new inequalities for Hermite-Hadamard divergence in information theory. *Stochastic Analysis and Applications.* Y. J. Cho, J. K. Kim and Y. K. Choi (Eds.), Nova Science Publishers, 2003, pp. 7-20.

[4] N. S. BARNETT, P. CERONE, S. S. DRAGOMIR and A. SOFO, Approximating Csiszár f-divergence by the use of Taylor's formula with integral remainder, *Math. Ineq. & Appl.*, **5** (3) (2002), 417-434.

[5] N. S. BARNETT, P. CERONE, S. S. DRAGOMIR and J. ROUMELIOTIS, Approximating Csiszár f-divergence via two integral identities and applications. *Stochastic Analysis and Applications*. Vol. 2, 1–17, Nova Sci. Publ., Huntington, NY, 2002.

[6] N. S. BARNETT, P. CERONE, S. S. DRAGOMIR and A. SOFO, Approximating two mappings associated to Csiszár f-divergence via Taylor's expansion, *PanAmerican Math. J.,* **12** (4) (2002), 105-117.

[7] N. S. BARNETT, P. CERONE, S. S. DRAGOMIR, A. SOFO, Approximating Csiszár f-divergence via an Ostrowski type identity for n-time differentiable functions. *Stochastic analysis and applications.* Vol. 2, 19–30, Nova Sci. Publ., Huntington, NY, 2002.

[8] P. R. BEESACK and J.E. PEČARIĆ, On the Jessen's inequality for convex functions, *J. Math. Anal. Appl*., **110** (1985), 536-552.

[9] R. BERAN, Minimum Hellinger distance estimates for parametric models, *Ann. Statist.,* 5 (1977), 445-463.

[10] M. BETH BASSAT, f-entropies, probability of error and feature selection, *Inform. Control*, **39** (1978), 227-242.

[11] A. BHATTACHARYYA, On a measure of divergence between two statistical populations defined by their probability distributions, *Bull. Calcutta Math. Soc.,* **35** (1943), 99-109.

[12] I. BURBEA and C. R. RAO, On the convexity of some divergence measures based on entropy function, *IEEE Trans. Inf. Th.,* **28** (3) (1982), 489-495.

[13] P. CERONE, On an identity for the Chebychev functional and some ramifications, *J. Ineq. Pure & Appl. Math*., **3** (1) (2002), Art. 4.

[14] P. CERONE and S. S. DRAGOMIR, Trapezoidal type rules from an inequalities point of view, *Handbook of Analytic-Computational Methods in Applied Mathematics*, Editor: G. Anastassiou, CRC Press, N. Y., (2000), 65-134.

[15] P. CERONE and S. S. DRAGOMIR, Midpoint type rules from an inequalities point of view, *Handbook of Analytic-Computational Methods in Applied Mathematics*, Editor: G. Anastassiou, CRC Press, N. Y., (2000), 135-200.

[16] P. CERONE, S. S. DRAGOMIR, and J. ROUMELIOTIS, Some Ostrowski type inequalities for n-time differentiable mappings and applications, *Demonstratio Math.,* **32** (2) (1999), 697-712.

[17] P. CERONE, S.S. DRAGOMIR, J. ROUMELIOTIS, and J. ŠUNDE, A new generalisation of the trapezoid formula for n-time differentiable mappings and applications, *Demonstratio Math.,***33** (4) (2000), 719-736.

[18] P. CERONE and S. S. DRAGOMIR, A refinement of the Grüss inequality and applications. *Tamkang J. Math.* **38** (2007), no. 1, 37–49. Preprint RGMIA *Res. Rep. Coll.* **5** (2002), No. 2, Art. 14. [Online http://rgmia.org/v5n2.php].

[19] C.H. CHEN, *Statistical Pattern Recognition,* Rocelle Park, New York, Hoyderc Book Co., 1973.

[20] C.K. CHOW and C. N. LIN, Approximating discrete probability distributions with dependence trees, *IEEE Trans. Inf. Th.,* **14** (3) (1968), 462-467.

[21] T.M. COVER and J. A. THOMAS, *Elements of Information Theory*, John Wiley & Sons Inc., 1991.

[22] I. CSISZÁR, A note on Jensen's inequality, *Studia Sci. Math. Hung.,* **1** (1966) 185–188.

[23] I. CSISZÁR, Information–type measures of differences of probability distributions and indirect observation, *Studia Sci. Math. Hung.,* **2** (1967) 299–318.

[24] I. CSISZÁR, On topological properties of f-divergence, *Studia Sci. Math. Hung.,* **2** (1967) 329–339.

[25] I. CSISZÁR, Information measures: A critical survey, *Trans. 7th Prague Conf. on Info. Th., Statist. Decis. Funct., Random Processes and 8th European Meeting of Statist.*, Volume B, Academia Prague, 1978, 73-86.

[26] I. CSISZÁR and J. KÖRNER, *Information Theory: Coding Theorem for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[27] D. DACUNHA-CASTELLE, *Ecole d'été de probabilités de Saint–Fleur*, VII-1977, Berlin, Heidelberg, New York: Springer, 1978.

[28] S. S. DRAGOMIR, An improvement of Jensen's inequality, *Bull. Math. Soc. Sci. Math. Roumanie*, **34** (**82**) (1990), No. 4, 291-296.

[29] S. S. DRAGOMIR, Some refinements of Ky Fan's inequality, *J. Math. Anal. Appl.*, **163** (2) (1992), 317-321.

[30] S. S. DRAGOMIR, Some refinements of Jensen's inequality, *J. Math. Anal. Appl.*, **168** (2) (1992), 518-522.

[31] S. S. DRAGOMIR, A further improvement of Jensen's inequality, *Tamkang J. Math.*, **25** (1) (1994), 29-36.

[32] S. S. DRAGOMIR, A new improvement of Jensen's inequality, *Indian J. Pure and Appl. Math.*, **26** (10) (1995), 959-968.

[33] S. S. DRAGOMIR, New estimation of the remainder in Taylor's formula using Grüss' type inequalities and applications, *Math. Ineq. and Appl.*, **2** (2) (1999), 183-193.

[34] S. S. Dragomir, On the Ostrowski's integral inequality for mappings of bounded variation and applications, *RGMIA Res. Rep. Coll.* **2** (1999) No. 1, 73–80.

[35] S. S. DRAGOMIR, An improvement of the remainder estimate in the generalised Taylor's formula, *RGMIA Res. Rep. Coll.*, **3** (2000), No. 1, Article 1.

[36] S. S. DRAGOMIR, Other inequalities for Csiszár divergence and applications, Preprint, RGMIA Monographs, Victoria University, 2000.

[37] S.S. DRAGOMIR, Some inequalities for (m, M)-convex mappings and applications for the Csiszár f-divergence in Information Theory, *Math. J. Ibaraki Univ.*, 33 (2001), 35-50.

[38] S. S. DRAGOMIR, Some inequalities for two Csiszár divergences and applications, *Mat. Bilten* (Macedonia), **25** (2001), 73-90.

[39] S. S. DRAGOMIR, On a reverse of Jessen's inequality for isotonic linear functionals, *J. Ineq. Pure & Appl. Math.*, **2**(2001), No. 3, Article 36.

[40] S. S. DRAGOMIR, An upper bound for the Csiszár f-divergence in terms of variational distance and applicaitons, *PanAmerican Math. J.*, **12**(4) (2002), 165-173.

[41] S. S. DRAGOMIR, Upper and lower bounds for Csiszár f-divergence in terms of Hellinger discrimination and applications, *Nonlinear Anal. Forum* (Korea), **7**(1) (2002), 1-13.

[42] S. S. DRAGOMIR, Bounds for f-divergences under likelihood ratio constraints, *Appl. of Math.* (Czech Republic), **48**(3) (2003), 205-223.

[43] S.S. DRAGOMIR, Some inequalities for the Csiszár f-divergence, *J. KSIAM* (Korea), **7**(1) (2003), 63-77.

[44] S. S. DRAGOMIR, New inequalities for Csiszár divergence and applications. *Acta Math. Vietnam.* **28** (2003), no. 2, 123–134.

[45] S. S. DRAGOMIR, A Grüss type inequality for isotonic linear functionals and applications. *Demonstratio Math.* **36** (2003), no. 3, 551–562. Preprint RGMIA *Res. Rep. Coll.* **5**(2002), Suplement, Art. 12. [Online http://rgmia.org

[46] S. S. DRAGOMIR, A converse inequality for the Csiszár Φ-divergence. *Tamsui Oxf. J. Math. Sci.* **20** (2004), no. 1, 35–53. Preprint in S. S. Dragomir (Ed.): *Inequalities for Csiszár f-Divergence in Information Theory,* RGMIA Monographs, Victoria University, 2000, [http://rgmia.org/monographs/csiszar_list.html#chap1]./v5(E).php].

[47] S. S. DRAGOMIR, Some inequalities for the Csiszár f-divergence when f is an L-Lipschitzian function and applications, *Italian J. Pure & Appl. Math.,* **15** (2004), 57-76.

[48] S. S. DRAGOMIR, *Discrete Inequalities of the Cauchy-Bunyakovsky-Schwarz Type*, Nova Science Publishers, NY, 2004.

[49] S. S. DRAGOMIR, *Semi-inner Products and Applications*, Nova Science Publishers Inc., NY, 2004.

[50] S. S. DRAGOMIR, Bounds for the normalized Jensen functional, *Bull. Austral. Math. Soc.* **74**(3)(2006), 471-476.

[51] S. S. DRAGOMIR, A generalisation of Cerone's identity and applications, *Oxford Tamsui J. Math. Sci.* **23** (2007), No. 1, 79-90. Preprint *RGMIA Res. Rep. Coll.*, **8**(2) (2005), Art. 19.

[52] S. S. DRAGOMIR, Inequalities for Stieltjes integrals with convex integrators and applications, *Appl. Math. Lett.*, **20** (2007), 123-130.

[53] S. S. DRAGOMIR, Bounds for the deviation of a function from the chord generated by its extremities. *Bull. Aust. Math. Soc.* **78** (2008), no. 2, 225–248.

[54] S. S. DRAGOMIR, A refinement of Jensen's inequality with applications for f-divergence measures, *Taiwanese J. Math.* **14** (2010), no. 1, 153–164.

[55] S. S. DRAGOMIR, A new refinement of Jensen's inequality in linear spaces with applications. *Math. Comput. Modelling* **52** (2010), no. 9-10, 1497–1505.

[56] S. S. DRAGOMIR, A refinement and a divided difference reverse of Jensen's inequality with applications, Preprint *RGMIA Res. Rep. Coll.* **14** (2011), Art. 74.

[57] S. S. DRAGOMIR, Inequalities in terms of the Gâteaux derivatives for convex functions on linear spaces with applications. *Bull. Aust. Math. Soc.* **83** (2011), no. 3, 500–517.

[58] S. S. DRAGOMIR, Some Slater's type inequalities for convex functions defined on linear spaces and applications. *Abstr. Appl. Anal.* **2012**, Art. ID 168405, 16 pp.

[59] S. S. DRAGOMIR, Some reverses of the Jensen inequality with applications. *Bull. Aust. Math. Soc.* **87** (2013), no. 2, 177–194.

[60] S. S. DRAGOMIR, Reverses of the Jensen inequality in terms of first derivative and applications. *Acta Math. Vietnam.* **38** (2013), no. 3, 429–446

[61] S. S. DRAGOMIR, N. M. DRAGOMIR and K. PRANESH, Some Estimations of Kraft number and related results, *Soochow Journal of Mathematics*, **24** (4), 291-296.

[62] S. S. DRAGOMIR and N. M. IONESCU, Some converse of Jensen's inequality and applications. *Rev. Anal. Numér. Théor. Approx.* **23** (1994), no. 1, 71–78.

[63] S. S. DRAGOMIR and V. GLUŠČEVIĆ, New estimates for the Kullback-Leibler distance and applications, in *Inequality Theory and Applications, Vol. 1*, Y. J. CHo, J. K. Kim and S. S. Dragomir (Eds.), Nova Science Publishers, New York, 2001, pp. 123-138.

[64] S. S. DRAGOMIR and V. GLUŠČEVIĆ, Approximating Csiszár $f-$divergence via a generalised Taylor formula, in *Inequalities for Csiszár $f-$divergence in Information Theory*, S.S. Dragomir (Ed.), RGMIA Monographs, Victoria University, 2001.

[65] S. S. DRAGOMIR, V. GLUŠČEVIĆ and C. E. M. PEARCE, Approximations for Csiszár f-divergence via midpoint inequalities. *Inequality theory and applications.* Vol. I, 139–153, Nova Sci. Publ., Huntington, NY, 2001.

[66] S. S. DRAGOMIR, V. GLUŠČEVIĆ and C. E. M. PEARCE, The approximation of Csiszár f-divergence for absolutely continuous mappings. *Inequality theory and applications*. Vol. I, 155–170, Nova Sci. Publ., Huntington, NY, 2001.

[67] S. S. DRAGOMIR, V. GLUŠČEVIĆ and C. E. M. PEARCE The approximation of Csiszár f-divergence for absolutely continuous mappings. *Inequality theory and applications*. Vol. I, 155–170, Nova Sci. Publ., Huntington, NY, 2001.

[68] S. S. DRAGOMIR, V. GLUŠČEVIĆ and C. E. M. PEARCE, The approximation of Csiszár f-divergence for mappings of bounded variation. *Inequality theory and applications.* Vol. I, 171–181, Nova Sci. Publ., Huntington, NY, 2001.

[69] S. S. DRAGOMIR, V. GLUŠČEVIĆ and C. E. M. PEARCE, Csiszár f-divergence, Ostrowski's inequality and mutual information. *Nonlinear Anal.* **47** (2001), no. 4, 2375–2386.

[70] S. S. DRAGOMIR and C. J. GOH, A counterpart of Jensen's discrete inequality for differentiable convex mappings and applications in information theory, *Math. Comput. Modelling,* **24**(2) (1996) 1–11.

[71] S. S. DRAGOMIR and C. J. GOH, Some bounds on entropy measures in information theory, *Appl. Math. Lett.,* **10** (3) (1997) 23–28.

[72] S. S. DRAGOMIR and C. J. GOH, Some bounds on entropy measures in information theory, *Appl. Math. Lett.,* **10** (1997), 23-28.

[73] S. S. DRAGOMIR and C. J. GOH, Some counterpart inequalities in for a functional associated with Jensen's inequality, *J. of Ineq. & Appl.,* **1** (1997), 311-325.

[74] S. S. DRAGOMIR and C. J. GOH, A counterpart of Jensen's continuous inequality and applications in information theory, *Anal. St. Univ. "Al. I. Cuza", Iaşi*, **XLVII** (2001), 239-262.

[75] S. S. DRAGOMIR, J. PEČARIĆ and L. E. PERSSON, Properties of some functionals related to Jensen's inequality, *Acta Math. Hung.*, **70**(1-2) (1996), 129-143.

[76] S. S. DRAGOMIR, M. SHOLZ and J. SUNDE, Some upper bounds for relative entropy and applications, *Computer and Math. with Appl.*, Vol. **39** (2000), No. 9-10, 91-100.

[77] S. S. DRAGOMIR and S. WANG, Application of Ostrowski's inequality to the estimation of error bounds for some special means and for some numerical quadrature rules, *Appl. Math. Lett.*, **11** (1998), 105–109.

[78] N. M. DRAGOMIR and S. S. DRAGOMIR, An inequality for logarithms and its applications in coding theory, *Indian Journal of Mathematics*, **43** (1) (2001), 13–20.

[79] B. R. FRIEDEN, Image enhancement and restoration, *Picture Processing and Digital Filtering* (T. S. Huang, Editor), Springer-Verlag, Berlin, 1975.

[80] D. V. GOKHALE and S. KULLBACK, *Information in Contingency Tables*, New York, Merul Dekker, 1978.

[81] R. G. GALLAGER, *Information Theory and Reliable Communications,* J. Wiley, New York, 1968.

[82] J. H. HAVRDA and F. CHARVAT, Quantification method classification process: concept of structural α-entropy, *Kybernetika*, **3** (1967), 30-35.

[83] E. HELLINGER, Neue Bergrüirdung du Theorie quadratisher Formerus von uneudlichvieleu Veränderlicher, *J. für reine and Augeur. Math.*, **36** (1909), 210-271.

[84] H. JEFFREYS, An invariant form for the prior probability in estimating problems, *Proc. Roy. Soc. London*, **186** A (1946), 453-461.

[85] J. H. JUSTICE (editor), *Maximum Entropy and Bayssian Methods in Applied Statistics*, Cambridge University Press, Cambridge, 1986.

[86] T. T. KADOTA and L. A. SHEPP, On the best finite set of linear observables for discriminating two Gaussian signals, *IEEE Trans. Inf. Th.*, **13** (1967), 288-294.

[87] J. N. KAPUR, On the roles of maximum entropy and minimum discrimination information principles in Statistics, *Technical Address of the 38th Annual Conference of the Indian Society of Agricultural Statistics*, 1984, 1-44.

[88] J. N. KAPUR, A comparative assessment of various measures of directed divergence, *Advances in Management Studies*, **3** (1984), 1-16.

[89] T. KAILATH, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans. Comm. Technology.*, Vol COM-15 (1967), 52-60.

[90] D. KAZAKOS and T. COTSIDAS, A decision theory approach to the approximation of discrete probability densities, *IEEE Trans. Perform. Anal. Machine Intell.*, **1** (1980), 61-67.

[91] J. H. B. KEMPERMAN, On the optimum note of transmitting information, *Ann. Math. Statist.*, **40** (1969), 2158-2177.

[92] C. KRAFT, Some conditions of consistency and uniform consistency of statistical procedure, *Univ. of California Publ. in Statistics, Ph.D. Thesis*, **1** (1955), 125–142.

[93] M. G. KREIN and A. A. NNUDEL'MAN, *The Markov moment problem and extremal problems*, Transl. Math. Monographs **50**, Amer. Math. Soc., 1977.

[94] S. KULLBACK, *Information Theory and Statistics*, J. Wiley, New York, 1959.

[95] S. KULLBACK, A lower bound for discrimination information in terms of variation, *IEEE Trans. Inf. Th.*, **13** (1967), 126-127.

[96] S. KULLBACK, Correction to a lower bound for discrimination information in terms of variation, *IEEE Trans. Inf. Th.*, **16** (1970), 771-773.

[97] S. KULLBACK and R. A. LEIBLER, On information and sufficiency, *Ann. Math. Stat.*, **22** (1951), 79-86.

[98] R.M. LEAHY and C.E. GOUTIS, An optimal technique for constraint-based image restoration and mensuration, *IEEE Trans. on Acoustics, Speech and Signal Processing*, **34** (1986), 1692-1642.

[99] L. LECAM, *Asymptotic Methods in Statistical Decision Theory*, New York: Springer, 1986.

[100] F. LIESE and I. VAJDA, *Convex Statistical Distances*, Teubner Verlag, Leipzig, 1987.

[101] J. LIN, Divergence measures based on the Shannon entropy, *IEEE Trans. Inf. Th.*, **37** (1) (1991), 145-151.

[102] J. LIN and S. K. M. WONG, A new directed divergence measure and its characterization, *Int. J. General Systems*, **17** (1990), 73-81.

[103] M. MATIĆ, *Jensen's Inequality and Applications in Information Theory* (in Croatian), Ph.D. Thesis, Univ. of Zagreb, 1999.

[104] M. MATIĆ, C. E. M. PEARCE and J. PEČARIĆ, Improvements of some bounds on entropy measures in information theory, *Math. Inequal. and Appl.*, 1 (1998), 295–304.

[105] M. MATIĆ, C. E. M. PEARCE and J. PEČARIĆ, Refinements of some bounds in information theory, *J. Austr. Math. Soc. Ser. B*, to appear.

[106] M. MATIĆ, J. PEČARIĆ and N. UJEVIĆ, On new estimation of the remainder in generalised Taylor's formula, *Math. Ineq. and Appl.*, **2** (3) (1999), 343-361.

[107] H. P. McKEAN, JR., Speed of approach to equilibrium for Koc's caricature of a Maximilian gas, *Arch. Ration. Mech. Anal.*, **21** (1966), 343-367..

[108] M. MEI, The theory of genetic distance and evaluation of human races, *Japan J. Human Genetics*, **23** (1978), 341-369

[109] D. S. MITRINOVIĆ, *Analytic Inequalities*, Springer Verlag, 1970.

[110] D. S. MITRINOVIĆ, J. E. PEČARIĆ and A.M. FINK, *Classical and New Inequalities in Analysis*, Kluwer Academic Publishers, 1993.

[111] D. S. MITRINOVIĆ, J. E. PEČARIĆ and A. M. FINK, *Inequalities for functions and their integrals and derivatives*, Kluwer Academic, Dordrecht, 1994.

[112] J. PEČARIĆ and S. S. DRAGOMIR, A refinements of Jensen inequality and applications, *Studia Univ. Babeş-Bolyai, Mathematica*, **24** (1) (1989), 15-19.

[113] J. E. PEČARIĆ, F. PROSCHAN and Y. L. TONG, *Convex Functions, Partial Orderings and Statistical Applications*, Academic Press, Inc., 1992.

[114] E.C. PIELOU, *Ecological Diversity*, Wiley, New York, 1975.

[115] M. S. PINSKER, Information and Information Stability of Random variables and processes, (in Russian), Moscow: Izv. Akad. Nouk, 1960.

[116] C. R. RAO, Diversity and dissimilarity coefficients: a unified approach, *Theoretic Population Biology*, **21** (1982), 24-43.

[117] A. RÉNYI, On measures of entropy and information, *Proc. Fourth Berkeley Symp. Math. Stat. and Prob., University of California Press*, **1** (1961), 547-561.

[118] A. SEN, *On Economic Inequality*, Oxford University Press, London 1973.

[119] C. E. SHANNON, A mathematical theory of communication, *Bull. Sept. Tech. J.*, **27** *(1948), 370-423 and 623-656.*

[120] B. D. SHARMA and D. P. MITTAL, New non-additive measures of relative information, *Journ. Comb. Inf. Sys. Sci.*, **2** (4)(1977), 122-132.

[121] H. SHIOYA and T. DA-TE, A generalisation of Lin divergence and the derivative of a new information divergence, *Elec. and Comm. in Japan*, **78** (7) (1995), 37-40.

[122] S. SIMIĆ, On a global upper bound for Jensen's inequality, *J. Math. Anal. Appl.* **343**(2008), 414-419.

[123] M. S. SLATER, A companion inequality to Jensen's inequality, *J. Approx. Theory*, **32**(1981), 160-166.

[124] I. J. TANEJA, *Generalised Information Measures and their Applications* (`http://www.mtm.ufsc.br/~taneja/bhtml/bhtml.html`).

[125] H. THEIL, *Economics and Information Theory*, North-Holland, Amsterdam, 1967.

[126] H. THEIL, *Statistical Decomposition Analysis*, North-Holland, Amsterdam, 1972.

[127] F. TOPSØE, Some inequalities for information divergence and related measures of discrimination, *RGMIA Res. Rep. Coll.* **2** (1999) No 1, 85–98.

[128] G. T. TOUSSAINT, Sharper lower bounds for discrimination in terms of variation, *IEEE Trans. Inf. Th.*, **21** (1975), 99-100.

[129] I. VAJDA, Note on discrimination information and variation, *IEEE Trans. Inf. Th.*, **16** (1970), 771-773.

[130] I. VAJDA, *Theory of Statistical Inference and Information,* Dordrecht-Boston, Kluwer Academic Publishers, 1989.

[131] V. A. VOLKONSKI and J. A. ROZANOV, Some limit theorems for random function -I, (English Trans.), *Theory Prob. Appl., (*USSR), **4** (1959), 178-197.

5. ANEX - PAPERS ON f-DIVERGENCE

A search on the database *MathSciNet* with the key word *"f-divergence"* produces the following list of papers so far.

(1) Csiszár, I. On topology properties of f-divergences. Studia Sci. Math. Hungar. 2 1967 329–339.
(2) Vajda, I. On the f-divergence and singularity of probability measures. Collection of articles dedicated to the memory of Alfréd Rényi, I. Period. Math. Hungar. 2 (1972), 223–234.
(3) Rosen, Joe Necessity and sufficiency of f being a divergence for vanishing of variational derivatives of f: general case. Lett. Math. Phys. 3 (1979), no. 3, 237–240.
(4) Abellanas, L.; Galindo, A.; Martínez Alonso, L. On a paper by J. Rosen: "Necessity and sufficiency of f being a divergence for vanishing of variational derivatives of f: general case" [Lett. Math. Phys. 3 (1979), no. 3, 237–240; MR 80d:49003]. Lett. Math. Phys. 3 (1979), no. 6, 453.
(5) Mussmann, D. Sufficiency and f-divergences. Studia Sci. Math. Hungar. 14 (1979), no. 1-3, 37–41 (1982).
(6) Perez, Albert Discrimination rate least favorable pairs of distributions for φ-contaminated statistical hypotheses or with f-divergence like neighborhoods. Transactions of the ninth Prague conference on information theory, statistical decision functions, random processes, Vol. A (Prague, 1982), 53–66, Reidel, Dordrecht, 1983.
(7) Pardo, L.; Menéndez, M. L.; Pardo, J. A. The f*-divergence as a criterion of comparison between fuzzy information systems. Kybernetes 15 (1986), no. 3, 189–194.
(8) Pardo, Leandro The measure of f-divergence as a stopping rule in the sequential random sampling in a Bayesian context. Statistica (Bologna) 46 (1986), no. 2, 243–251.
(9) Pardo Llorente, J. A.; Menéndez Calleja, Ma. Luisa f*-divergence measures in the Bayesian context. (Spanish) Proceedings of the XIIth Portuguese-Spanish Conference on Mathematics, Vol. III (Portuguese) (Braga, 1987), 275–281, Univ. Minho, Braga, 1987.
(10) Menéndez, Ma. Luisa The f*-divergences in the sequential observation of fuzzy information systems. Kybernetes 17 (1988), no. 4, 41–52.
(11) Feldman, D.; Österreicher, F. A note on f-divergences. Studia Sci. Math. Hungar. 24 (1989), no. 2-3, 191–200.
(12) Vos, Paul W. Geometry of f-divergence. Ann. Inst. Statist. Math. 43 (1991), no. 3, 515–537.
(13) Kafka, P.; Österreicher, F.; Vincze, I. On powers of f-divergences defining a distance. Studia Sci. Math. Hungar. 26 (1991), no. 4, 415–422.
(14) Vos, Paul W. Minimum f-divergence estimators and quasi-likelihood functions. Ann. Inst. Statist. Math. 44 (1992), no. 2, 261–279.

(15) Pardo, J. A.; Menéndez, M. L.; Pardo, L. f*-divergence as a Bayes criterion for the comparison of experiments. (Spanish) Stochastica 13 (1992), no. 1, 75–87.
(16) Kateri, Maria; Papaioannou, Takis f-divergence association models. Internat. J. Math. Statist. Sci. 3 (1994), no. 2, 179–203 (1995).
(17) Dragomir, S. S.; Gluščević, V.; Pearce, C. E. M. Csiszár f-divergence, Ostrowski's inequality and mutual information. Proceedings of the Third World Congress of Nonlinear Analysts, Part 4 (Catania, 2000). Nonlinear Anal. 47 (2001), no. 4, 2375–2386.
(18) Dragomir, S. S.; Gluščević, V.; Pearce, C. E. M. Approximations for Csiszár f-divergence via midpoint inequalities. Inequality theory and applications. Vol. I, 139–153, Nova Sci. Publ., Huntington, NY, 2001.
(19) Dragomir, S. S.; Gluščević, V.; Pearce, C. E. M. The approximation of Csiszár f-divergence for absolutely continuous mappings. Inequality theory and applications. Vol. I, 155–170, Nova Sci. Publ., Huntington, NY, 2001.
(20) Dragomir, S. S.; Gluščević, V.; Pearce, C. E. M. The approximation of Csiszár f-divergence for mappings of bounded variation. Inequality theory and applications. Vol. I, 171–181, Nova Sci. Publ., Huntington, NY, 2001.
(21) Barnett, N. S.; Cerone, P.; Dragomir, S. S.; Sofo, A. Approximating Csiszár f-divergence by the use of Taylor's formula with integral remainder. Inequalities, 2001 (Timişoara). Math. Inequal. Appl. 5 (2002), no. 3, 417–434.
(22) Dragomir, S. S. An upper bound for the Csiszár f-divergence in terms of the variational distance and applications. PanAmer. Math. J. 12 (2002), no. 4, 43–54.
(23) Barnett, N. S.; Cerone, P.; Dragomir, S. S.; Sofo, A. Approximating two mappings associated to Csiszár f-divergence via Taylor's expansion. PanAmer. Math. J. 12 (2002), no. 4, 105–117.
(24) Dragomir, S. S. Upper and lower bounds for Csiszár f-divergence in terms of Hellinger discrimination and applications. Nonlinear Anal. Forum 7 (2002), no. 1, 1–13.
(25) Barnett, N. S.; Cerone, P.; Dragomir, S. S.; Roumeliotis, J. Approximating Csiszár f-divergence via two integral identities and applications. Stochastic analysis and applications. Vol. 2, 1–17, Nova Sci. Publ., Huntington, NY, 2002.
(26) Barnett, N. S.; Cerone, P.; Dragomir, S. S.; Sofo, A. Approximating Csiszár f-divergence via an Ostrowski type identity for n-time differentiable functions. Stochastic analysis and applications. Vol. 2, 19–30, Nova Sci. Publ., Huntington, NY, 2002.
(27) Dragomir, S. S. Bounds for f-divergences under likelihood ratio constraints. Appl. Math. 48 (2003), no. 3, 205–223.
(28) Dragomir, S. S. A generalized f-divergence for probability vectors and applications. PanAmer. Math. J. 13 (2003), no. 4, 61–69.
(29) Taneja, Inder Jeet; Kumar, Pranesh Relative information of type s, Csiszár's f-divergence, and information inequalities. Inform. Sci. 166 (2004), no. 1-4, 105–125.
(30) Cerone, Pietro; Dragomir, Sever S.; Österreicher, Ferdinand Bounds on extended f-divergences for a variety of classes. Kybernetika (Prague) 40 (2004), no. 6, 745–756.
(31) Anastassiou, George A. The Csiszar's f-divergence as a measure of dependence. An. Univ. Oradea Fasc. Mat. 11 (2004), 27–50.
(32) Anastassiou, George A. Higher order optimal approximation of Csiszar's f-divergence. Nonlinear Anal. 61 (2005), no. 3, 309–339
(33) Kumar, P.; Chhina, S. A symmetric information divergence measure of the Csiszár's f-divergence class and its bounds. Comput. Math. Appl. 49 (2005), no. 4, 575–588.
(34) Anastassiou, George Basic optimal approximation of Csiszar's f-divergence. Approximation theory XI: Gatlinburg 2004, 15–23, Mod. Methods Math., Nashboro Press, Brentwood, TN, 2005.

(35) Cerone, P.; Dragomir, S. S. Approximation of the integral mean divergence and f-divergence via mean results. Math. Comput. Modelling 42 (2005), no. 1-2, 207–219.
(36) Anastassiou, George A. Fractional and other approximation of Csiszar's f-divergence. Rend. Circ. Mat. Palermo (2) Suppl. No. 76 (2005), 197–212.
(37) Anastassiou, George A. Representations and estimates to Csiszar's f-divergence. PanAmer. Math. J. 16 (2006), no. 1, 83–106.
(38) Gushchin, Alexander A.; Zhdanov, Denis A. A minimax result for f-divergences. From stochastic calculus to mathematical finance, 287–294, Springer, Berlin, 2006.
(39) Gilardoni, Gustavo L. On the minimum f-divergence for given total variation. C. R. Math. Acad. Sci. Paris 343 (2006), no. 11-12, 763–766.
(40) Anastassiou, George A. Optimal approximation of discrete Csiszar f-divergence. Math. Balkanica (N.S.) 20 (2006), no. 2, 185–206.
(41) Baig, M. A. K.; Dar, Rayees Ahmad A new generalized bound for Csiszar f-divergence in terms of Renyi's entropy of order α and its applications. Antarct. J. Math. 4 (2007), no. 1, 65–77
(42) Jain, K. C.; Srivastava, Amit On symmetric information divergence measures of Csiszar's f-divergence class. Far East J. Appl. Math. 29 (2007), no. 3, 477–491.
(43) Jain, K. C.; Srivastava, Amit On some new symmetric information divergence measures of Csiszar's f-divergence class. J. Rajasthan Acad. Phys. Sci. 6 (2007), no. 4, 449–456.
(44) Baig, M. A. K.; Dar, Rayees Ahmad Upper and lower bounds for Csiszar f-divergence in terms of symmetric J-divergence and applications. Indian J. Math. 49 (2007), no. 3, 263–278.
(45) Liese, Friedrich; Vajda, Igor f-divergences: sufficiency, deficiency and testing of hypotheses. Advances in inequalities from probability theory and statistics, 113–149, Adv. Math. Inequal. Ser., Nova Sci. Publ., New York, 2008.
(46) Jain, K. C.; Srivastava, Amit On some new symmetric information divergence measures of Csizar's f-divergence class & information inequalities. Antarct. J. Math. 5 (2008), no. 2, 87–105.
(47) Gushchin, A. A. On an extension of the concept of f-divergence. (Russian) ; translated from Teor. Veroyatn. Primen. 52 (2007), no. 3, 468–489 Theory Probab. Appl. 52 (2008), no. 3, 439–455.
(48) Nguyen, XuanLong; Wainwright, Martin J.; Jordan, Michael I. On surrogate loss functions and f-divergences. Ann. Statist. 37 (2009), no. 2, 876–904.
(49) Anastassiou, George A. Riemann-Liouville and Caputo fractional approximation of Csiszar's f-divergence. Sarajevo J. Math. 5(17) (2009), no. 1, 3–12.
(50) Amari, Shun-Ichi α-divergence is unique, belonging to both f-divergence and Bregman divergence classes. IEEE Trans. Inform. Theory 55 (2009), no. 11, 4925–4931.
(51) Gilardoni, Gustavo L. Corrigendum to the note "On the minimum f-divergence for given total variation" [C. R. Acad. Sci. Paris, Ser. I 343 (2006) 763–766] [MR2284708]. C. R. Math. Acad. Sci. Paris 348 (2010), no. 5-6, 299.
(52) Dragomir, S. S. A refinement of Jensen's inequality with applications for f-divergence measures. Taiwanese J. Math. 14 (2010), no. 1, 153–164.
(53) Petz, Dénes From f-divergence to quantum quasi-entropies and their use. Entropy 12 (2010), no. 3, 304–325.
(54) Jain, K. C.; Mathur, Ruchi An information divergence measure of Csiszar's f-divergence class and its bounds. J. Rajasthan Acad. Phys. Sci. 9 (2010), no. 1, 65–76.

(55) Dragomir, Sever S.; Koumandos, Stamatis Some inequalities for f-divergence measures generated by 2n-convex functions. Acta Sci. Math. (Szeged) 76 (2010), no. 1-2, 71–86.
(56) Qiao, Yu; Minematsu, Nobuaki A study on invariance of f-divergence and its application to speech recognition. IEEE Trans. Signal Process. 58 (2010), no. 7, 3884–3890.
(57) Gilardoni, Gustavo L. On Pinsker's and Vajda's type inequalities for Csiszár's f-divergences. IEEE Trans. Inform. Theory 56 (2010), no. 11, 5377–5386.
(58) Guntuboyina, Adityanand Lower bounds for the minimax risk using f-divergences, and applications. IEEE Trans. Inform. Theory 57 (2011), no. 4, 2386–2399.
(59) Harremoës, Peter; Vajda, Igor On pairs of f-divergences and their joint range. IEEE Trans. Inform. Theory 57 (2011), no. 6, 3230–3235.
(60) Hiai, Fumio; Mosonyi, Milán; Petz, Dénes; Bény, Cédric Quantum f-divergences and error correction. Rev. Math. Phys. 23 (2011), no. 7, 691–747.
(61) Jain, K. C.; Saraswat, Ram Naresh Some bounds of Csiszar's F-divergence measure in terms of the well known divergence measures of information theory. Int. J. Math. Sci. Eng. Appl. 5 (2011), no. 5, 1–11.
(62) Kanamori, Takafumi; Suzuki, Taiji; Sugiyama, Masashi f-divergence estimation and two-sample homogeneity test under semiparametric density-ratio models. IEEE Trans. Inform. Theory 58 (2012), no. 2, 708–720.
(63) Srivastava, Amit Some new information inequalities involving f-divergences. Cybern. Inf. Technol. 12 (2012), no. 2, 3–10.
(64) Chen, Guang-Sheng; Shi, Xing-Jun Generalizations of Hölder inequalities for Csiszar's f-divergence. J. Inequal. Appl. 2013, 2013:151, 6 pp.
(65) Bhatia, P. K.; Singh, Surender On a new Csiszar's f-divergence measure. Cybern. Inf. Technol. 13 (2013), no. 2, 43–57.
(66) Molnár, Lajos; Nagy, Gergő; Szokol, Patrícia Maps on density operators preserving quantum f-divergences. Quantum Inf. Process. 12 (2013), no. 7, 2309–2323.
(67) Cawston, Suzanne; Vostrikova, Lioudmila Levy preservation and associated properties for f-divergence minimal equivalent martingale measures. Prokhorov and contemporary probability theory, 163–196, Springer Proc. Math. Stat., 33, Springer, Heidelberg, 2013.
(68) Werner, Elisabeth M. f-divergence for convex bodies. Asymptotic geometric analysis, 381–395, Fields Inst. Commun., 68, Springer, New York, 2013.
(69) Dragomir, S. S. A generalization of f-divergence measure to convex functions defined on linear spaces. Commun. Math. Anal. 15 (2013), no. 2, 1–14.
(70) Barthe, Gilles; Olmedo, Federico Beyond differential privacy: composition theorems and relational logic for f-divergences between probabilistic programs. Automata, languages, and programming. Part II, 49–60, Lecture Notes in Comput. Sci., 7966, Springer, Heidelberg, 2013.
(71) Moslehian, Mohammad Sal; Kian, Mohsen Non-commutative f-divergence functional. Math. Nachr. 286 (2013), no. 14-15, 1514–1529.
(72) Ren, Feng Ying; Li, Xing Si Coherent risk measure based on f-divergence. (Chinese) J. Dalian Univ. Technol. 53 (2013), no. 5, 772–776.
(73) Cawston, S.; Vostrikova, L. f-divergence minimal equivalent martingale measures and optimal portfolios for exponential Lévy models with a change-point. Seminar on Stochastic Analysis, Random Fields and Applications VII, 305–336, Progr. Probab., 67, Birkhäuser/Springer, Basel, 2013.
(74) Guntuboyina, Adityanand; Saha, Sujayam; Schiebinger, Geoffrey Sharp inequalities for f-divergences. IEEE Trans. Inform. Theory 60 (2014), no. 1, 104–121.

(75) Cawston, S.; Vostrikova, L. An f-divergence approach for optimal portfolios in exponential Lévy models. Inspired by finance, 83–101, Springer, Cham, 2014.
(76) Dragomir, Silvestru Sever A new quantum f-divergence for trace class operators in Hilbert spaces. Entropy 16 (2014), no. 11, 5853–5875.
(77) Rastegin, Alexey E. On quantum conditional entropies defined in terms of the f-divergences. Rep. Math. Phys. 73 (2014), no. 3, 393–411.
(78) Saigusa, Yusuke; Tahata, Kouji; Tomizawa, Sadao Orthogonal decomposition of symmetry model using the ordinal quasi-symmetry model based on f-divergence for square contingency tables. Statist. Probab. Lett. 101 (2015), 33–37.
(79) Caglar, Umut; Werner, Elisabeth M. Mixed f-divergence and inequalities for log-concave functions. Proc. Lond. Math. Soc. (3) 110 (2015), no. 2, 271–290.
(80) Dragomir, Silvestru Sever Some inequalities for f-divergences via Slater's inequality for convex functions. J. Inequal. Spec. Funct. 6 (2015), no. 2, 5–14.
(81) Cerone, Pietro; Dragomir, Sever S.; Kikianty, Eder Jensen-Ostrowski type inequalities and applications for f-divergence measures. Appl. Math. Comput. 266 (2015), 304–315.
(82) Jain, Kamal C.; Chhabra, Praphull New information inequalities on new generalized f-divergence and applications. Matematiche (Catania) 70 (2015), no. 2, 271–281.
(83) Dragomir, S. S. A survey of reverse inequalities for f-divergence measure in information theory. Computation, cryptography, and network security, 209–251, Springer, Cham, 2015.
(84) Niezgoda, Marek Vector joint majorization and generalization of Csiszár-Körner's inequality for f-divergence. Discrete Appl. Math. 198 (2016), 195–205.
(85) Kian, Mohsen; Dragomir, S. S. f-divergence functional of operator log-convex functions. Linear Multilinear Algebra 64 (2016), no. 2, 123–135.
(86) Torra, Vicenç; Narukawa, Yasuo; Sugeno, Michio On the f-divergence for non-additive measures. Fuzzy Sets and Systems 292 (2016), 364–379.
(87) Virosztek, Dániel Quantum f-divergence preserving maps on positive semidefinite operators acting on finite dimensional Hilbert spaces. Linear Algebra Appl. 501 (2016), 242–253.
(88) de Wet, T.; Österreicher, F. Fisher's information and the class of f-divergences based on extended Arimoto's entropies. South African Statist. J. 50 (2016), no. 1, 43–64.
(89) Kluza, Paweł; Niezgoda, Marek Generalizations of Crooks and Lin's results on Jeffreys-Csiszár and Jensen-Csiszár f-divergences. Phys. A 463 (2016), 383–393.
(90) Gushchin, A. A.; The minimum increment of f-divergences given total variation distances. Math. Methods Statist. 25 (2016), no. 4, 304–312.