Kybernetika

Pietro Cerone; Sever Silvestru Dragomir; Ferdinand Österreicher
Bound on extended f-divergences for a variety of classes

BOUNDS ON EXTENDED f–DIVERGENCES FOR A VARIETY OF CLASSES

PIETRO CERONE, SEVER S. DRAGOMIR AND FERDINAND ÖSTERREICHER

The concept of f-divergences was introduced by Csiszár in 1963 as measures of the 'hardness' of a testing problem depending on a convex real valued function f on the interval $[0, \infty)$. The choice of this parameter f can be adjusted so as to match the needs for specific applications. The definition and some of the most basic properties of f-divergences are given and the class of χ^α-divergences is presented. Ostrowski's inequality and a Trapezoid inequality are utilized in order to prove bounds for an extension of the set of f-divergences. The class of χ^α-divergences and four further classes of f-divergences are used in order to investigate limitations and strengths of the inequalities derived.

Keywords: f-divergences, bounds, Ostrowski's inequality

AMS Subject Classification: 94A17

1. INTRODUCTION TO f–DIVERGENCES

Let $(X, \mathcal{A})$ be a measurable space, satisfying $|\mathcal{A}| > 2$ and μ be a σ-finite measure on $(X, \mathcal{A})$ and let $\mathcal{P}$ be the set of all probability measures on $(X, \mathcal{A})$ which are absolutely continuous with respect to μ. For $P, Q \in \mathcal{P}$ let $p = \frac{\mathrm{d}P}{\mathrm{d}\mu}$ and $q = \frac{\mathrm{d}Q}{\mu}$ denote the Radon–Nikodym derivatives of P and Q with respect to μ. Two probability measures $P, Q \in \mathcal{P}$ are called *orthogonal* $(Q \perp P)$ if $P(\{q = 0\}) = Q(\{p = 0\}) = 1$.

Furthermore, let $\mathcal{F}$ be the set of convex functions $f : [0, \infty) \mapsto (-\infty, \infty]$ which are finite on $(0, \infty)$ and continuous at 0 (i.e. $f(0) = \lim_{u \downarrow 0} f(u)$), $\mathcal{F}_0 = \{f \in \mathcal{F} : f(1) = 0\}$ and let $D_- f$ and $D_+ f$ denote the *left-hand side derivative* and the *right-hand side derivative* of f respectively. Further, if $f \in \mathcal{F}$ then f^* defined by

$$
f^*(u) = \begin{cases} u f\left(\frac{1}{u}\right) & \text{for} \quad u \in (0, \infty) \\ \lim_{v \to \infty} f(v)/v & \text{for} \quad u = 0 \end{cases}
$$

is also in $\mathcal{F}$ and is called the $*$-*conjugate* (convex) function of f.

Definition 1.1. (Csiszár [2]) Let $f \in \mathcal{F}$ and $P, Q \in \mathcal{P}$. Then

$$I_f(Q, P) = \int_X p \cdot f\left(\frac{q}{p}\right) d\mu$$

is called the *f-divergence* of the probability distributions Q and P.

The following two theorems contain the most basic properties of f-divergences. For their proof we refer the reader to, for example, Chapter 1 of [6].

Theorem 1.2. (Uniqueness and Symmetry Theorem) Let $f, f_1 \in \mathcal{F}$ and f^* be the *-conjugate of f. Then

$$I_{f_1}(Q, P) = I_f(Q, P) \quad \forall \, (P, Q) \in \mathcal{P}^2 \quad \text{iff} \quad \exists c \in \mathbb{R} : f_1(u) - f(u) = c(u - 1)$$

and $I_{f^*}(Q, P) = I_f(Q, P) \, \forall \, (P, Q) \in \mathcal{P}^2$ iff

(i) $\exists c \in \mathbb{R} : f^*(u) - f(u) = c(u - 1)$.

Theorem 1.3. (Range of Values Theorem) Let $f \in \mathcal{F}$. Then

$$f(1) \le I_f(Q, P) \le f(0) + f^*(0) \quad \forall \, Q, P \in \mathcal{P}.$$

In the first inequality, equality holds if $Q = P$. Further, this equality holds if and only if $Q = P$, provided that

(ii) f is strictly convex at 1.

In the second, equality holds if $Q \perp P$. Further, this equality holds if and only if $Q \perp P$, provided that

(iii) $f(0) + f^*(0) < \infty$.

Remark 1.4. In order for an f-divergence to be nonnegative, it is necessary to restrict oneself to the class $\mathcal{F}_0$. Since the values of an f-divergence are not changed when replacing the function $f \in \mathcal{F}_0$ by $f(u) - c(u - 1)$, it is convenient to take a value $c \in [D_- f(1), D_+ f(1)]$ in order to get a nonnegative function f.

Remark 1.5. If $f \in \mathcal{F}_0$, then

$$I_f(Q, P) = f(0) \cdot P(\{q = 0\}) + f^*(0) \cdot Q(\{p = 0\}) + \int_{\{p \cdot q > 0\}} p \cdot f\left(\frac{q}{p}\right) d\mu$$

holds, where $P(\{q = 0\})$ is the amount of singularity of the distribution P with respect to Q and $Q(\{p = 0\})$ is the amount of singularity of the distribution Q with respect to P. Therefore $f(0) = \infty$ or $f^*(0) = \infty$ imply $I_f(Q, P) = \infty$ unless $P \ll Q$ (i.e. $P(\{q = 0\}) = 0$) or $Q \ll P$ (i.e. $Q(\{p = 0\}) = 0$) respectively.

In the Sections 1 and 3 we consider five classes of f-divergences, from which the classes (I), (II) and (III) are well-known. All elements f of the five classes satisfy

$f(1) = 0$, $f(u) \geq 0 \; \forall u \in [0, \infty)$ and property (ii). The elements of the classes (III), (IV) and (V) satisfy, in addition, (i) and (iii), and thus complete the basic requirements in order to allow for a metric divergence. We conclude this section by presenting the first class, which we need in Theorem 2.1, the main result of this paper.

(I) χ^α-divergences. The f-divergences of this class, which is generated by the functions χ^α, $\alpha \in [1, \infty)$, defined by

$$\chi^\alpha(u) = |u - 1|^\alpha \; , \; u \in [0, \infty) \, ,$$

have the form

$$I_f(Q, P) = \int_X p \left| \frac{q}{p} - 1 \right|^\alpha \mathrm{d}\mu = \int_X p^{1-\alpha} \, |q - p|^\alpha \, \mathrm{d}\mu \, ,$$

whereby, in view of Remark 1.5, we assume $Q \ll P$ provided $\alpha > 1$.

From this class only the parameter $\alpha = 1$ provides a distance, namely the total variation distance $V(Q, P) = \int_X |q - p| \, \mathrm{d}\mu$. The most prominent special case of this class is, however, Karl Pearson's χ^2-divergence.

The following result is a refinement of the second inequality in Theorem 1.3.

Theorem 1.6. Let $f \in \mathcal{F}_0$ satisfy condition (iii) and let $\tilde{f} = (f + f^*)/2$. Then

$$I_f(Q, P) \leq \tilde{f}(0) I_{\chi^1}(Q, P) \quad \forall \, Q, \, P \in \mathcal{P}. \tag{1.1}$$

P r o o f. The proof relies on the following simple fact

$$f(u) \leq \begin{cases} f(0)(1 - u) & \text{for} \quad u \in [0, 1] \\ f^*(0)(u - 1) & \text{for} \quad u \in (1, \infty). \end{cases} \tag{1.2}$$

First we consider the set $\{p \cdot q > 0\}$. Setting $u = \frac{q}{p}$, multiplying the resulting inequality by $p > 0$ and integrating over the set $\{p \cdot q > 0\}$ yields

$$\int_{\{p \cdot q > 0\}} p \cdot f\left(\frac{q}{p}\right) \mathrm{d}\mu \; \leq \; f(0) \left(P(\{p \geq q > 0\}) - Q(\{p \geq q\})\right)$$
$$+ f^*(0) \left(Q(\{q > p > 0\}) - P(\{q > p\})\right).$$

Together with the part for the remaining set $\{p = 0\} \cup \{q = 0\}$, covered by Remark 1.5, this yields

$$\begin{aligned} I_f(Q, P) \; &\leq \; f(0) \left(P(\{p \geq q\}) - Q(\{p \geq q\})\right) \\ &\quad + f^*(0) \left(Q(\{q > p\}) - P(\{q > p\})\right) \\ &= \; \tilde{f}(0) V(Q, P), \end{aligned}$$

where the latter is easily seen by splitting up $X = \{p \geq q\} \cup \{q > p\}$ and applying $\int_X q \mathrm{d}\mu = \int_X p \mathrm{d}\mu = 1$. $\qquad \square$

Remark 1.7. For functions $f \in \mathcal{F}_0$ which satisfy, in addition to (iii) also condition (i), $\bar{f}(0) = f(0)$ holds. The most simple example satisfying all conditions (i), (ii) and (iii) is χ^1 which corresponds to the total variation distance. Note that for functions $f \in \mathcal{F}_0$ given by equality in (1.2) equality holds in (1.1) as well.

2. GENERAL INEQUALITIES FOR EXTENDED f-DIVERGENCES

In the following result, the notion of the f-divergence is extended to the class $\bar{\mathcal{F}}$ of functions $f : [0, \infty) \to \mathbb{R}$ which are locally absolutely continuous functions on $[0, \infty)$ the derivative f' of which is of bounded variation on each compact interval $[1, u]$ (or $[u, 1]$). In this context $\mathsf{V}_a^b(g)$ denotes the variation of a function g over an interval with the endpoints a and b. Corresponding to the set $\mathcal{F}_0$ in Section 1 let, further, $\bar{\mathcal{F}}_0 = \{f \in \bar{\mathcal{F}} : f(1) = 0\}$.

Theorem 2.1. Let $f \in \bar{\mathcal{F}}$ be such that the derivative f' satisfies

$$\mathsf{V}_1^u(f') \leq V \, |u - 1|^{\rho - 1} \quad \text{for any } u \in (0, \infty), \tag{2.1}$$

where the constants $V > 0$ and $\rho \in (1, \infty)$ are given.

Then for any $P, Q \in \mathcal{P}$, with $Q \ll P$, one has the inequalities

$$|I_f(Q, P) - f(1)| \leq V I_{\chi^\rho}(Q, P) \tag{2.2}$$

and

$$\left| I_f(Q, P) - f(1) - I_{h_f^{(i)}}(Q, P) \right| \leq \frac{V}{2} I_{\chi^\rho}(Q, P) \tag{2.3}$$

where the functions $h_f^{(i)}$, $i \in \{1, 2\}$ are defined by

$$h_f^{(i)}(u) = \begin{cases} (u - 1)f'\left(\frac{u+1}{2}\right) & \text{for } i = 1 \\ \frac{1}{2}(u - 1)f'(u) & \text{for } i = 2, \end{cases} \quad u \in (0, \infty) \tag{2.4}$$

and $I_{\chi^\rho}(Q, P)$ is defined as in Section 1.

Proof. In the first part we prove the inequalities (2.2) and (2.3) for the case $i = 1$. We use the following inequality obtained in [3]. If $g : [a, b] \to \mathbb{R}$ is of bounded variation over $[a, b]$ then for any $y \in [a, b]$

$$\left| g(y) - \frac{1}{b - a} \int_a^b g(s) \, ds \right| \leq \left[\frac{1}{2} + \left| \frac{y - \frac{a+b}{2}}{b - a} \right| \right] \mathsf{V}_a^b(g). \tag{2.5}$$

Now, the application of the inequality (2.5) for the choices for the choices $g = f'$, $a = \min(1, u)$, $b = \max(1, u)$, $u \in (0, \infty)$ and $y = 1$ yields in view of (2.2)

$$|f(\dot{u}) - f(1) - (u - 1)f'(1)| \leq |u - 1| \, \mathsf{V}_1^u(f') \leq V \, |u - 1|^\rho \tag{2.6}$$

for any $u \in (0, \infty)$.

Choosing $u = q/p$ in (2.6) we get, after multiplying by $p \geq 0$

$$\left| pf\left(\frac{q}{p}\right) - f(1)p - (q - p)f'(1) \right| \leq Vp^{1-\rho} |q - p|^{\rho} .$$

Integrating this over X, using the triangle inequality and taking into account the fact that $\int_X p \, d\mu = \int_X q \, d\mu = 1$, we deduce the desired inequality (2.2).

From the inequality (2.5), we also get by choosing $g = f'$, $a = \min(1, u)$, $b = \max(1, u)$, $u \in (0, \infty)$ and $y = \frac{u+1}{2}$

$$\left| f(u) - f(1) - (u - 1)f'\left(\frac{u + 1}{2}\right) \right| \leq \frac{1}{2} |u - 1| \, V_1^u(f') \leq \frac{V}{2} |u - 1|^{\rho} \tag{2.7}$$

for any $u \in (0, \infty)$.

If we choose $u = \frac{q}{p}$ in (2.7) we get, by multiplying with $p \geq 0$,

$$\left| pf\left(\frac{q}{p}\right) - f(1)p - (q - p)f'\left(\frac{q + p}{2p}\right) \right| \leq \frac{V}{2} p^{1-\rho} |q - p|^{\rho} ,$$

which gives, by integration over X, the desired inequality (2.3) for the case $i = 1$.

In the following second part we prove the inequality (2.3) for the case $i = 2$. For this part we use the following inequality obtained in [1]. If $g : [a, b] \to \mathbb{R}$ is of bounded variation over $[a, b]$ then for any $y \in [a, b]$

$$\left| \frac{(b - y)g(b) + (y - a)g(a)}{b - a} - \frac{1}{b - a} \int_a^b g(s) \, ds \right| \leq \left[\frac{1}{2} + \left| \frac{y - \frac{a+b}{2}}{b - a} \right| \right] V_a^b(g) . \tag{2.8}$$

The application of the inequality (2.8) for the choices $g = f'$, $a = \min(1, u)$, $b = \max(1, u)$, $u \in (0, \infty)$ and $y = \frac{u+1}{2}$ yields

$$\left| f(u) - f(1) - (u - 1)\frac{f'(1) + f'(u)}{2} \right| \leq \frac{1}{2} |u - 1| \, V_1^u(f') \leq \frac{V}{2} |u - 1|^{\rho} \tag{2.9}$$

for any $u \in (0, \infty)$.

Choosing, in (2.9), $u = \frac{q}{p}$ and by multiplying with $p \geq 0$ and then integrating over X, we deduce the desired inequality (2.3) for the case $i = 2$. $\qquad \square$

Remark 2.2. The constant $\frac{1}{2}$ both in (2.5) and (2.8) is best possible in the sense that it cannot be replaced by a smaller one. Dragomir et al. [4] utilised Ostrowski's inequality to examine bounds for extended f-divergences, however, they assumed stricter conditions on the functions requiring absolute continuity.

Remark 2.3. As mentioned in Section 1, f-divergences frequently use (convex) functions $f \in \mathcal{F}_0$. For these and, more generally, for $f \in \bar{\mathcal{F}}_0$ the term $f(1)$ in the inequalities (2.2) and (2.3) is to be dropped.

Remark 2.4. For $f \in \mathcal{F}$ we interpret $f'(u)$ for those (at most denumerably many) points $u \in (0, \infty)$ for which $D_- f(u) < D_+ f(u)$ as $f'(u) = \frac{1}{2} (D_- f(u) + D_+ f(u))$. In this way we can extend (2.1) – and hence Theorem 2.1 – to the case $\rho = 1$. By considering the function $\chi^1(u) = |u - 1|$ we see, in view of $\left(\chi^1\right)'(u) = \text{signum}(u - 1)$, that (2.1) has the form $V_1^u \left(\left(\chi^1\right)'\right) = 1$ for any $u \in (0, \infty) \backslash \{1\}$ and, consequently, that (2.6) gives the form

$$\chi^1(u) = V |u - 1|^\rho$$

with $V = 1$ and $\rho = 1$. This shows that (2.2) is sharp in the case of the total variation distance. It is not difficult to see when restricting the application of Theorem 2.1 to functions $f \in \mathcal{F}$ that (2.6) – and consequently (2.2) – is sharp only for functions of the form

$$f(u) = c |u - 1| + d. \tag{2.10}$$

(By the way, the sign-function can be used to show that the constant $\frac{1}{2}$ in (2.6) is best possible.) Finally we note that (2.1) – and consequently (2.2) – is in the case of the power $\rho = 1$ limited to functions of the form (2.10) whereas inequality (1.1) applies to all functions $f \in \mathcal{F}_0$ satisfying condition (iii).

3. INVESTIGATION OF THE RESULTS

In this section we use the class of χ^α-divergences and four further classes of f-divergences in order to investigate both applicability and quality of the inequalities (2.2) and (2.3) derived in Section 2. Note that the functions f of all these classes belong to $\mathcal{F}_0$ and satisfy (ii) and $f(u) \geq 0$ for any $u \in [0, \infty)$. The investigation of the quality is mostly done by comparing the bound V achieved in (2.2) with the best possible bound for the given power $\rho \geq 1$. The latter is the smallest constant c which satisfies

$$f(u) \leq c |u - 1|^\rho \quad \forall u \in [0, \infty)$$

for the given $f \in \mathcal{F}_0$. Note that $c \geq f(0)$ with strict inequality provided that $f(0) > |D_+ f(0)| / \rho$.

When applying Theorem 2.1 we typically encounter two types of numerical losses. The first arising from the inequality (2.1) and the second by either the application of Ostrowski's inequality (2.5) or by the application of the Trapezoid inequality (2.8).

Note that inequality (2.1) can only be applied to the derivative of a (convex) function $f \in \mathcal{F}$ provided that

$$|D_+ f(0)| < \infty.$$

Although this condition is necessary but not sufficient for (2.1) the (finite) value $|D_+ f(0)|$ turns out to be crucial since it provides the constant V in (2.1) for all cases investigated, except in class (V) for the parameters $\alpha \in (2, \infty)$.

Remark 3.1. The functions $h_f^{(i)}$, $i \in \{1, 2\}$, defined in (2.4) satisfy

$$h_f^{(i)}(1) = 0 \quad \text{and} \quad \left(h_f^{(i)}\right)'(1) = \begin{cases} f'(1) & \text{for} \quad i = 1 \\ \frac{1}{2} f'(1) & \text{for} \quad i = 2. \end{cases}$$

Further, for functions $f \in \mathcal{F}$ with continuous derivatives f' on $(0, \infty)$ with the properties (ii) and $f'(1) = 0$ it holds

$$\left(h_f^{(i)}\right)'(u) = \begin{cases} < 0 & \text{for} \quad u < 1 \\ = 0 & \text{for} \quad u = 1 \\ > 0 & \text{for} \quad u > 1 \end{cases} \quad \text{and hence } h_f^{(i)}(u) \geq 0 \,\, \forall u \in (0, \infty), \,\, i \in \{1, 2\}.$$

Functions $f \in \mathcal{F}_0$ with continuous second derivative satisfying $f''(1) > 0$ yield, in addition,

$$\left(h_f^{(i)}\right)''(1) = \begin{cases} f''(1) & \text{for} \quad i = 1 \\ 2f''(1) & \text{for} \quad i = 2 \end{cases} > 0.$$

Therefore, the corresponding divergences $I_{h_f^{(i)}}(Q, P)$ are disparities as defined in Mendéndez, Morales, Pardo and Vajda [7]. For further applications of related disparities we refer to Wegenkittl [10] and [11].

Note, however, that the functions $h_f^{(i)}$ do not necessarily inherit the property of convexity from the functions f, so that $I_{h_f^{(i)}}$ are in general not f-divergences in the strict sense.

From the following classes of f-divergences the classes (I), (II) and (III) are well-known. A detailed discussion can be found for example in Liese and Vajda [6], Chapter 2. The classes (IV) and (V) were introduced respectively by Puri and Vincze [9] and Österreicher and Vajda [8].

Our intention in the sequel is to give concise statements of our findings. Therefore we omit any of (the partially laborious) details.

(I) χ^α-Divergences
For this class

$$\left| (\chi^\alpha)'(u) \right| = \alpha \left| u - 1 \right|^{\alpha - 1} \begin{cases} \forall\, u \in (0, \infty) \backslash \{1\} & \text{if} \quad \alpha \in [1, 2) \\ \forall\, u \in (0, \infty) & \text{if} \quad \alpha \in [2, \infty) \end{cases}$$

holds. Therefore inequality (2.1) – and hence (2.2) and (2.3) – apply for all parameters $\alpha \in [1, \infty)$. The corresponding powers and constants are $\rho = V = \alpha$. In this case the inequality (2.2) has the trivial form

$$I_{\chi^\alpha}(Q, P) \leq \alpha I_{\chi^\alpha}(Q, P).$$

Obviously, for the parameter $\alpha = 1$, which corresponds to the total variation distance, equality holds. Note that although in (2.1) equality holds true in this case a

considerable loss is caused by the application of the inequality (2.7) for the parameters $\alpha > 1$. For this class of f-divergences for which the functions

$$h_{\chi^\alpha}^{(i)} = \begin{cases} \frac{\alpha}{2^{\alpha-1}}\chi^\alpha & \text{for} \quad i = 1 \\ \frac{\alpha}{2}\chi^\alpha & \text{for} \quad i = 2 \end{cases}$$

are elements of the set $\mathcal{F}_0$ of convex functions, the inequalities (2.3) are of a similar quality as (2.2).

Concerning the discussion of the total variation distance, which appears as a special case also in the classes (III) and (IV) for the parameter $\alpha = 1$ and in class (V) for the parameter $\alpha = \infty$, we refer to Remark 2.4 and the discussion in class (I).

(II) Power Divergences

From this class, generated by the functions

$$f_\alpha(u) = \begin{cases} u - 1 - \ln u & \text{for} \quad \alpha = 0 \\ \frac{\alpha u + 1 - \alpha - u^\alpha}{\alpha(1-\alpha)} & \text{for} \quad \alpha \in \mathbb{R} \setminus \{0, 1\}, \quad u \in [0, \infty), \\ 1 - u + u \ln u & \text{for} \quad \alpha = 1 \end{cases}$$

only the parameter $\alpha = \frac{1}{2}$ $\left(f_{1/2}(u)/2 = (\sqrt{u} - 1)^2\right)$ allows for defining a distance, namely the Hellinger Distance

$$H(Q, P) = \left[\int_X (\sqrt{q} - \sqrt{p})^2 \, d\mu\right]^{1/2}.$$

The functions of this class satisfy

$$|D_+ f_\alpha(0)| = \begin{cases} \frac{1}{\alpha-1} & \text{for} \quad \alpha \in (1, \infty) \\ \infty & \text{for} \quad \alpha \in (-\infty, 1]. \end{cases}$$

It can be shown that, in addition, f_α' fails to satisfy (2.1) for any power ρ for the parameters $\alpha \in (2, \infty)$.

For the parameters $\alpha \in (1, 2]$, however, f_α' satisfies (2.1) – and hence (2.2) and (2.3) – with $\rho = 2$ and $V = \frac{1}{\alpha-1}$. In this case the inequality (2.2) has the form

$$I_{f_\alpha}(Q, P) \le c_\alpha I_{\chi^2}(Q, P), \tag{3.1}$$

with $c_\alpha = \frac{1}{\alpha-1}$. Note that (3.1) holds for all $\alpha \in (0, 2]$ and that $c_\alpha = f_\alpha(0) = \frac{1}{\alpha}$ is best possible. The functions

$$h_{f_\alpha}^{(i)}(u) = \begin{cases} \frac{u-1}{\alpha-1}\left(\left(\frac{u+1}{2}\right)^{\alpha-1} - 1\right) & \text{for} \quad i = 1 \\ \frac{(u-1)(u^{\alpha-1}-1)}{2(\alpha-1)} & \text{for} \quad i = 2 \end{cases}$$

in the inequality (2.3) turn out to be elements of $\mathcal{F}_0$, so that

$$I_{h_{f_\alpha}^{(1)}}(Q,P) = \frac{1}{\alpha - 1} \int_X (q - p) \left(\left(\frac{q + p}{2p} \right)^{\alpha - 1} - 1 \right) d\mu, \quad \alpha \in (1, 2]$$

and

$$I_{h_{f_\alpha}^{(2)}}(Q,P) = \frac{1}{2(\alpha - 1)} \int_X (q - p) \left(\left(\frac{q}{p} \right)^{\alpha - 1} - 1 \right) d\mu \quad \alpha \in (1, 2]$$

are f-divergences in the strict sense.

(III) Matusita's Divergences

The elements of this class, which is generated by the functions φ_α, $\alpha \in (0, 1]$, given by

$$\varphi_\alpha(u) = |1 - u^\alpha|^{1/\alpha}, \quad u \in [0, \infty),$$

are prototypes of metric divergences, providing the distances $[I_{\varphi_\alpha}(Q,P)]^\alpha$.

These functions satisfy $|D_+\varphi_\alpha(0)| = \infty$ for all $\alpha \in (0, 1)$. Therefore the results of Section 2 cannot be applied.

Note, however, that the functions $\alpha \mapsto \varphi_\alpha(u)$ are strictly monotone increasing for every $u \in (0, \infty) \backslash \{1\}$ and hence

$$I_{\varphi_\alpha}(Q,P) \leq I_{\varphi_\beta}(Q,P) \leq I_{\varphi_1}(Q,P) = I_{\chi^1}(Q,P) \quad \text{for all } 0 < \alpha < \beta < 1.$$

For the latter inequality compare also with Theorem 1.6, Remark 1.7 and Remark 2.3.

(IV) Puri–Vincze Divergences

This class is generated by the functions Φ_α, $\alpha \in [1, \infty)$ given by

$$\Phi_\alpha(u) = \frac{|u - 1|^\alpha}{2(u + 1)^{\alpha - 1}}, \quad u \in [0, \infty).$$

As shown in [5] this class provides the distances $[I_{\Phi_\alpha}(Q,P)]^{1/\alpha}$.

These functions satisfy

$$|\Phi'_\alpha(u)| = \frac{|u - 1|^{\alpha - 1}(2\alpha + (u - 1))}{2(u + 1)^\alpha} \quad \forall u \in (0, \infty)$$

and $|D_+\Phi_\alpha(0)| = \alpha - \frac{1}{2}$. Further, it can be shown that

$$|\Phi'_\alpha(u)| \leq \left(\alpha - \frac{1}{2} \right) \times |u - 1|^{\alpha - 1} \quad \forall u \in (0, \infty).$$

Therefore Theorem 2.1 can be applied for all parameters $\alpha \in (1, \infty)$ with $\rho = \alpha$ and $V = \alpha - \frac{1}{2}$.

The inequality (2.2) takes the form

$$I_{\Phi_\alpha}(Q,P) \le c_\alpha I_{\chi^\alpha}(Q,P) \tag{3.2}$$

with $c_\alpha = \alpha - \frac{1}{2}$ whereas, obviously, $c_\alpha = \Phi_\alpha(0) = \frac{1}{2}$ is best possible. Note also that none of the functions

$$h_{\Phi_\alpha}^{(i)}(u) = \begin{cases} \frac{|u-1|^\alpha(4\alpha+u-1)}{2(u+3)^\alpha} & \text{for} \quad i=1 \\ \frac{|u-1|^\alpha(2\alpha+u-1)}{4(u+1)^\alpha} & \text{for} \quad i=2, \end{cases} \quad u \in (0,\infty),$$

which are used in inequality (2.3), have a nonnegative second derivative, so that $I_{h_{\Phi_\alpha}^{(i)}}(Q,P)$ are disparities in the sense of Menéndez et al. [7] but not f-divergences in the strict sense.

We finally note that for the parameters $\alpha > 2$ the inequality (2.1) – and consequently (2.2) and (2.3) – holds not only for the power $\rho = \alpha$ but for all powers $\rho \in [2,\alpha]$.

(V) Divergences of Arimoto-type

This class is generated by the functions

$$\psi_\alpha(u) = \begin{cases} \frac{1}{1-1/\alpha}\left[(1+u^\alpha)^{1/\alpha} - 2^{1/\alpha-1}(1+u)\right] & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\} \\ (1+u)\ln(2) + u\ln(u) - (1+u)\ln(1+u) & \text{for} \quad \alpha = 1 \\ |1-u|/2 & \text{for} \quad \alpha = \infty. \end{cases}$$

As shown in [8] this class provides the distances $[I_{f_\alpha}(Q,P)]^{\min(\alpha,1/2)}$ for $\alpha \in (0,\infty)$ and $V(Q,P)/2$ for $\alpha = \infty$.

These functions satisfy

$$\psi_\alpha'(u) = \begin{cases} \frac{1}{1-1/\alpha}\left[(1+u^\alpha)^{1/\alpha-1}u^{\alpha-1} - 2^{1/\alpha-1}\right] & \text{for} \quad \alpha \in (0,\infty)\backslash\{1\} \\ \ln(2) + \ln(u) - \ln(u+1) & \text{for} \quad \alpha = 1, \end{cases}$$

$$\psi_\alpha''(u) = \alpha(1+u^\alpha)^{1/\alpha-2}u^{\alpha-2} \quad \text{and} \quad |D^+\psi_\alpha(0)| = \begin{cases} \infty & \text{for} \quad \alpha \in (0,1] \\ \frac{2^{1/\alpha-1}}{1-1/\alpha} & \text{for} \quad \alpha \in (1,\infty). \end{cases}$$

It can be shown that ψ_α' satisfies (2.1) – and hence (2.2) and (2.3) – with $\rho = 2$, $V = \frac{2^{1/\alpha-1}}{1-1/\alpha}$ for $\alpha \in (1,2]$ and an intricate form of V for $\alpha \in (2,\infty)$. Therefore Theorem 2.1 can be applied for the parameters $\alpha \in (1,\infty)$.

The inequality (2.2) takes the form

$$I_{\psi_\alpha}(Q,P) \le c_\alpha I_{\chi^2}(Q,P) \tag{3.3}$$

with $c_\alpha = \frac{2^{1/\alpha - 1}}{1 - 1/\alpha}$ for $\alpha \in (1, 2]$. Note, however, that (3.3) holds for all $\alpha \in (0, \infty)$ and that

$$c_\alpha = \psi_\alpha(0) = \begin{cases} \frac{1 - 2^{1/\alpha - 1}}{1 - 1/\alpha} & \text{for} \quad \alpha \in (0, \infty) \backslash \{1\} \\ \ln(2) & \text{for} \quad \alpha = 1 \end{cases}$$

is best possible for all $\alpha \in (0, \alpha^*]$, $\alpha^* = \left(2 - \frac{\ln(3)}{\ln(2)} \right)^{-1} > 2$. In addition, it can be shown that the best possible c_α satisfies $c_\alpha \in (\psi_\alpha(0), d_\alpha)$ for all $\alpha \in (\alpha^*, \infty)$ where

$$d_\alpha = \max_{u \in (0, \infty)} \psi_\alpha''(u)/2 = (2\alpha - 1)^{1/\alpha - 2} (\alpha - 2)^{1 - 2/\alpha} (1 + \alpha)^{1 + 1/\alpha} /2, \quad \alpha > 2.$$

Finally we note that none of the functions

$$h_{\psi_\alpha}^{(i)}(u) = \begin{cases} \frac{u-1}{1 - 1/\alpha} \left[(2^\alpha + (u+1)^\alpha)^{1/\alpha - 1} (u+1)^{\alpha - 1} - 2^{1/\alpha - 1} \right] & \text{for} \quad i = 1 \\ \frac{u-1}{2(1 - 1/\alpha)} \left[(1 + u^\alpha)^{1/\alpha - 1} u^{\alpha - 1} - 2^{\frac{1}{\alpha} - 1} \right] & \text{for} \quad i = 2, \end{cases}$$

$u \in (0, \infty)$, which are used in inequality (2.3), have a nonnegative second derivative, so that $I_{h_{\psi_\alpha}^{(i)}}(Q, P)$ are disparities in the sense of Menéndez et al. [7] but not f-divergences in the strict sense.

ACKNOWLEDGEMENTS

REFERENCES

[1] P. Cerone, S. S. Dragomir, and C. E. M. Pearce: A generalized trapezoid inequality for functions of bounded variation. Turkish J. Math. *24* (2000), 2, 147–163.

[2] I. Csiszár: Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. Sci. *8* (1963), 85–107.

[3] S. S. Dragomir: On the Ostrowski's inequalities for mappings with bounded variation and applications. Math. Inequ. Appl. *4* (2001), 1, 59–66.

[4] S. S. Dragomir, V. Gluščević, and C. M. E. Pearce: Csiszár *f*-divergences, Ostrowski's inequality and mutual information. Nonlinear Anal. *47* (2001), 4, 2375–2386.

[5] P. Kafka, F. Österreicher, and I. Vincze: On powers of *f*-divergences defining a distance. Studia Sci. Math. Hungar. *26* (1991), 415–422.

[6] F. Liese and I. Vajda: Convex Statistical Distances. (Teubner–Texte zur Mathematik, Band 95.) Teubner, Leipzig 1987.

[7] M. Menéndez, D. Morales, L. Pardo, and I. Vajda: Two approaches to grouping of data and related disparity statistics. Comm. Statist. Theory Methods *27* (1998), 3, 609–633.

[8] F. Österreicher and I. Vajda: A new class of metric divergences on probability spaces and its applicability in statistics. Ann. Inst. Statist. Math. *55* (2003), 3, 639–653.

[9] M. L. Puri and I. Vincze: Measures of information and contiguity. Statist. Probab. Lett. *9* (1990), 223–228.

[10] S. Wegenkittl: Entropy estimators and serial tests for ergodic chains. IEEE Trans. Inform. Theory *47* (2001), 6, 2480–2489.

[11] S. Wegenkittl: A generalized ϕ-divergence for asymptotically multivariate normal models. J. Multivariate Anal. *83* (2002), 288–302.

Pietro Cerone and Sever S. Dragomir, School of Computer Science and Mathematics, Victoria University of Technology, P. O. Box 14428, Melbourne City MC, VIC 8001. Australia.
e-mails: pietro.cerone, sever.dragomir@vu.edu.au

Ferdinand Österreicher, Fachbereich Mathematik, Universität Salzburg, Hellbrunner-strasse 34, A-5020 Salzburg. Austria.
e-mail: ferdinand.oesterreicher@sbg.ac.at