Canad. Math. Bull. Vol. 60 (3), 2017 pp. 641–654
http://dx.doi.org/10.4153/CMB-2016-050-3

Mixed f-divergence for Multiple Pairs of Measures

Elisabeth Werner and Deping Ye

Abstract. In this paper, the concept of the classical f-divergence for a pair of measures is extended to the mixed f-divergence for multiple pairs of measures. The mixed f-divergence provides a way to measure the difference between multiple pairs of (probability) measures. Properties for the mixed f-divergence are established, such as permutation invariance and symmetry in distributions. An Alexandrov–Fenchel type inequality and an isoperimetric inequality for the mixed f-divergence are proved.

1 Introduction

In applications such as pattern matching, image analysis, statistical learning, and information theory, one often needs to compare two (probability) measures and to know whether they are similar to each other. Hence, finding the "right" quantity to measure the difference between two (probability) measures P and Q is central. Traditionally, people use the classical L_p distances between P and Q, such as the variational distance and the L_2 distance. However, the family of f-divergences is often more suitable to fulfill the goal than the classical L_p distance of measures.

The f-divergence $D_f(P,Q)$ of two probability measures P and Q was first introduced in [8] and independently in [2, 30] and was defined by

$$D_f(P,Q) = \int_X f\Big(\frac{p}{q}\Big)\, q\, d\mu. \tag{1.1}$$

Here, p and q are density functions of P and Q with respect to a measure μ on X. The idea behind the f-divergence is to replace, for instance, the function $f(t) = |t-1|$ in the variational distance by a general convex function f. Hence, the f-divergence includes various widely used divergences as special cases, such as, the variational distance, the Kullback–Leibler divergence [16], the Bhattacharyya distance [5], and many more. Consequently, the f-divergence receives considerable attention not only in the information theory (*e.g.,* [3, 7, 14, 17, 31]) but also in many other areas. We only mention convex geometry. Within the last few years, amazing connections have been discovered between notions and concepts from convex geometry and information theory, *e.g.,* [9, 10, 15, 24, 25, 32], leading to a totally new point of view and introducing a whole new set of tools in the area of convex geometry. In particular, it was observed in

Received by the editors June 1, 2016.
Published electronically September 8, 2016.
Partially supported by an NSF grant.
Supported by an NSERC grant and a start-up grant from Memorial University of Newfoundland.
AMS subject classification: 28-XX, 52-XX, 60-XX.
Keywords: Alexandrov–Fenchel inequality, f-dissimilarity, f-divergence, isoperimetric inequality.

[38] that one of the most important affine invariant notions, the L_p-affine surface area for convex bodies, *e.g.,* [18–20, 22, 34], is Rényi entropy from information theory and statistics. Rényi entropies are special cases of f-divergences and, consequently, were then introduced for convex bodies, and their corresponding entropy inequalities have been established in [39]. We also refer to [4], for more references related to the f-divergence.

Extension of the f-divergence from two (probability) measures to multiple (probability) measures is fundamental in many applications, such as statistical hypothesis test and classification to which much research has been devoted, for instance in [28, 29, 42]. Such extensions include, *e.g.,* the Matusita affinity [26, 27], the Toussaint affinity [37], the information radius [36] and the average divergence [35].

The $\mathbf{f}$-dissimilarity $D_{\mathbf{f}}(P_1, \dots, P_l)$ for (probability) measures $P_1, \dots, P_l$, introduced in [11, 12] for a convex function $\mathbf{f}: \mathbb{R}^l \to \mathbb{R}$, is a natural generalization of the f-divergence. It is defined as

$$D_{\mathbf{f}}(P_1, \dots, P_l) = \int_X \mathbf{f}(p_1, \dots, p_l)\, d\mu,$$

where the p_i's are density functions of the P_i's that are absolutely continuous with respect to μ. For a convex function f, the function $\mathbf{f}(x, y) = y f(x/y)$ is also convex on $x, y > 0$, and $D_{\mathbf{f}}(P, Q)$ is equal to the classical f-divergence defined in formula (1.1). Note that the Matusita affinity is related to

$$\mathbf{f}(x_1, \dots, x_l) = -\prod_{i=1}^{l} x_i^{1/l},$$

and the Toussaint affinity is related to $\mathbf{f}(x_1, \dots, x_l) = -\prod_{i=1}^{l} x_i^{a_i}$, where $a_i \geq 0$ and such that $\sum_{i=1}^{l} a_i = 1$.

Here, we introduce special $\mathbf{f}$-dissimilarities, namely the mixed f-divergence and the i-th mixed f-divergence, which can be viewed as vector forms of the usual f-divergence. We establish some basic properties of these quantities such as permutation invariance and symmetry in distributions. We prove an isoperimetric type inequality and an Alexandrov–Fenchel type inequality for the mixed f-divergence. Alexandrov–Fenchel inequality is a fundamental inequality in convex geometry, and many important inequalities such as the Brunn–Minkowski inequality and Minkowski's first inequality follow from it (see, *e.g.,* [9, 33]).

The paper is organized as follows. In Section 2 we establish some basic properties of the mixed f-divergence such as permutation invariance and symmetry in distributions. In Section 3 we prove the general Alexandrov–Fenchel inequality and isoperimetric inequality for the mixed f-divergence. Section 4 is dedicated to the i-th mixed f-divergence and its related isoperimetric type inequalities.

2 The Mixed f-Divergence

Throughout this paper, let (X, μ) be a finite measure space. For $1 \leq i \leq n$, let $P_i = p_i \mu$ and $Q_i = q_i \mu$ be probability measures on X that are absolutely continuous with respect to the measure μ. Moreover, we assume that for all $i = 1, \dots, n$, p_i and q_i are nonzero μ-a.e. We use $\vec{\mathbf{P}}$ and $\vec{\mathbf{Q}}$ to denote the vectors of probability measures, or, in short, probability vectors,

$$\vec{\mathbf{P}} = (P_1, P_2, \dots, P_n), \quad \vec{\mathbf{Q}} = (Q_1, Q_2, \dots, Q_n).$$

We use $\vec{p}$ and $\vec{q}$ to denote the vectors of density functions, or density vectors, for $\vec{\mathbf{P}}$ and $\vec{\mathbf{Q}}$ respectively,

$$\frac{d\vec{\mathbf{P}}}{d\mu} = \vec{p} = (p_1, p_2, \dots, p_n), \quad \frac{d\vec{\mathbf{Q}}}{d\mu} = \vec{q} = (q_1, q_2, \dots, q_n).$$

We make the convention that $0 \cdot \infty = 0$.

Denote by $\mathbb{R}^+ = \{x \in \mathbb{R} : x \geq 0\}$. Let $f\colon (0,\infty) \to \mathbb{R}^+$ be a non-negative convex or concave function. The $*$-adjoint function $f^*\colon (0,\infty) \to \mathbb{R}^+$ of f is defined by

$$f^*(t) = tf(1/t).$$

It is obvious that $(f^*)^* = f$ and that f^* is again convex (resp. concave) if f is convex (resp. concave).

Let $f_i\colon (0,\infty) \to \mathbb{R}^+$, $1 \leq i \leq n$, be either convex or concave functions. Denote by $\vec{\mathbf{f}} = (f_1, f_2, \dots, f_n)$ the vector of functions. We write

$$\vec{\mathbf{f}}^* = (f_1^*, f_2^*, \dots, f_n^*)$$

to be the $*$-adjoint vector for $\vec{\mathbf{f}}$.

Now we introduce *the mixed f-divergence* for $(\vec{\mathbf{f}}, \vec{\mathbf{P}}, \vec{\mathbf{Q}})$ as follows.

Definition 2.1 Let (X,μ) be a measure space. Let $\vec{\mathbf{P}}$ and $\vec{\mathbf{Q}}$ be two probability vectors on X with density vectors $\vec{p}$ and $\vec{q}$, respectively. The mixed f-divergence $D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}})$ for $(\vec{\mathbf{f}}, \vec{\mathbf{P}}, \vec{\mathbf{Q}})$ is defined by

$$D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}}) = \int_X \prod_{i=1}^n \Big[f_i\Big(\frac{p_i}{q_i}\Big) q_i \Big]^{\frac{1}{n}} d\mu. \tag{2.1}$$

Similarly, we define the mixed f-divergence for $(\vec{\mathbf{f}}, \vec{\mathbf{Q}}, \vec{\mathbf{P}})$ by

$$D_{\vec{\mathbf{f}}}(\vec{\mathbf{Q}}, \vec{\mathbf{P}}) = \int_X \prod_{i=1}^n \Big[f_i\Big(\frac{q_i}{p_i}\Big) p_i \Big]^{\frac{1}{n}} d\mu. \tag{2.2}$$

A special case is when all distributions P_i and Q_i are identical and equal to a probability distribution P. In this case,

$$D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}}) = D_{(f_1, f_2, \dots, f_n)}\big((P, P, \dots, P), (P, P, \dots, P)\big) = \prod_{i=1}^n \big[f_i(1) \big]^{\frac{1}{n}}.$$

Let $\pi \in S_n$ denote a permutation on $\{1, 2, \dots, n\}$ and denote

$$\pi(\vec{p}) = (p_{\pi(1)}, p_{\pi(2)}, \dots, p_{\pi(n)}).$$

One immediate result from Definition 2.1 is the following permutation invariance for $D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}})$.

Proposition 2.2 (Permutation invariance) *Let the vectors $\vec{\mathbf{f}}, \vec{\mathbf{P}}, \vec{\mathbf{Q}}$ be as above, and let $\pi \in S(n)$ be a permutation on $\{1, 2, \dots, n\}$. Then*

$$D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}}) = D_{\pi(\vec{\mathbf{f}})}\big(\pi(\vec{\mathbf{P}}), \pi(\vec{\mathbf{Q}})\big).$$

When all (f_i, P_i, Q_i) are equal to (f, P, Q), the mixed f-divergence is equal to the classical f-divergence, denoted by $D_f(P,Q)$, which takes the form

$$D_f(P,Q) = D_{(f,f,\dots,f)}\big((P,P,\dots,P),(Q,Q,\dots,Q)\big) = \int_X f\Big(\frac{p}{q}\Big) q\,d\mu.$$

As $f^*(t) = tf(1/t)$, one easily obtains a fundamental property for the classical f-divergence $D_f(P,Q)$, namely,

$$D_f(P,Q) = D_{f^*}(Q,P),$$

for all (f,P,Q). Similar results hold true for the mixed f-divergence. We show this now.

Let $0 \le k \le n$. We write $D_{\vec{\mathbf{f}},k}(\vec{\mathbf{P}},\vec{\mathbf{Q}})$ for

$$D_{\vec{\mathbf{f}},k}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = \int_X \prod_{i=1}^{k}\Big[f_i\Big(\frac{p_i}{q_i}\Big)q_i\Big]^{\frac{1}{n}} \times \prod_{i=k+1}^{n}\Big[f_i^*\Big(\frac{q_i}{p_i}\Big)p_i\Big]^{\frac{1}{n}}d\mu.$$

Clearly, $D_{\vec{\mathbf{f}},n}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}})$ and $D_{\vec{\mathbf{f}},0}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = D_{\vec{\mathbf{f}}^*}(\vec{\mathbf{Q}},\vec{\mathbf{P}})$, where

$$\vec{\mathbf{f}}^* = (f_1^*, f_2^*, \dots, f_n^*).$$

Then we have the following result for changing order of distributions.

Proposition 2.3 (Principle for changing order of distributions) *Let $\vec{\mathbf{f}},\vec{\mathbf{P}},\vec{\mathbf{Q}}$ be as above. Then, for any $0 \le k \le n$, one has $D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = D_{\vec{\mathbf{f}},k}(\vec{\mathbf{P}},\vec{\mathbf{Q}})$. In particular, $D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = D_{\vec{\mathbf{f}}^*}(\vec{\mathbf{Q}},\vec{\mathbf{P}})$.*

Proof Let $0 \le k \le n$. Then

$$\begin{aligned} D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) &= \int_X \prod_{i=1}^{k}\Big[f_i\Big(\frac{p_i}{q_i}\Big)q_i\Big]^{\frac{1}{n}} \times \prod_{i=k+1}^{n}\Big[f_i\Big(\frac{p_i}{q_i}\Big)q_i\Big]^{\frac{1}{n}}d\mu \\ &= \int_X \prod_{i=1}^{k}\Big[f_i\Big(\frac{p_i}{q_i}\Big)q_i\Big]^{\frac{1}{n}} \times \prod_{i=k+1}^{n}\Big[f_i^*\Big(\frac{q_i}{p_i}\Big)p_i\Big]^{\frac{1}{n}}d\mu = D_{\vec{\mathbf{f}},k}(\vec{\mathbf{P}},\vec{\mathbf{Q}}), \end{aligned}$$

where the second equality follows from $f_i(\frac{p_i}{q_i})q_i = f_i^*(\frac{q_i}{p_i})p_i$. ■

A direct consequence of Proposition 2.3 is the following symmetry principle for the mixed f-divergence.

Proposition 2.4 (Symmetry in distributions) *Let $\vec{\mathbf{f}},\vec{\mathbf{P}},\vec{\mathbf{Q}}$ be as above. Then $D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) + D_{\vec{\mathbf{f}}^*}(\vec{\mathbf{P}},\vec{\mathbf{Q}})$ is symmetric in $\vec{\mathbf{P}}$ and $\vec{\mathbf{Q}}$, namely,*

$$D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) + D_{\vec{\mathbf{f}}^*}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = D_{\vec{\mathbf{f}}}(\vec{\mathbf{Q}},\vec{\mathbf{P}}) + D_{\vec{\mathbf{f}}^*}(\vec{\mathbf{Q}},\vec{\mathbf{P}}).$$

Remark Proposition 2.3 says that $D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}})$ remains the same if one replaces any triple (f_i, P_i, Q_i) by (f_i^*, Q_i, P_i). It is also easy to see that, for all $0 \le k, l \le n$, one has

$$D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = D_{\vec{\mathbf{f}},k}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = D_{\vec{\mathbf{f}}^*,l}(\vec{\mathbf{Q}},\vec{\mathbf{P}}) = D_{\vec{\mathbf{f}}^*}(\vec{\mathbf{Q}},\vec{\mathbf{P}}).$$

Hence, for all $0 \le k, l \le n$,

$$D_{\vec{\mathbf{f}},k}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) + D_{\vec{\mathbf{f}}^*,l}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) + D_{\vec{\mathbf{f}}^*}(\vec{\mathbf{P}},\vec{\mathbf{Q}})$$

is symmetric in $\vec{\mathbf{P}}$ and $\vec{\mathbf{Q}}$.

Hereafter, we only consider the mixed f-divergence $D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}})$ defined in formula (2.1). Properties for the mixed f-divergence $D_{\vec{\mathbf{f}}}(\vec{\mathbf{Q}}, \vec{\mathbf{P}})$ defined in (2.2) follow along the same lines.

Now we list some important mixed f-divergences.

Examples (i) The total variation is a widely used f-divergence to measure the difference between two probability measures P and Q on (X, μ). It is related to function $f(t) = |t-1|$. Similarly, the *mixed total variation* is defined by

$$D_{TV}(\vec{\mathbf{P}}, \vec{\mathbf{Q}}) = \int_X \prod_{i=1}^n |p_i - q_i|^{\frac{1}{n}} \, d\mu.$$

It measures the difference between two probability vectors $\vec{\mathbf{P}}$ and $\vec{\mathbf{Q}}$.

(ii) For $a \in \mathbb{R}$, we denote by $a_+ = \max\{a, 0\}$. The *mixed relative entropy* or *mixed Kullback–Leibler divergence* of $\vec{\mathbf{P}}$ and $\vec{\mathbf{Q}}$ is defined by

$$D_{KL}\big(\vec{\mathbf{P}}, \vec{\mathbf{Q}}\big) = D_{(f_+, \ldots, f_+)}\big(\vec{\mathbf{P}}, \vec{\mathbf{Q}}\big) = \int_X \prod_{i=1}^n \Big[p_i \ln\Big(\frac{q_i}{p_i}\Big)\Big]_+^{\frac{1}{n}} d\mu,$$

where $f(t) = t \ln t$. When $P_i = P = p\mu$ and $Q_i = Q = q\mu$ for all $i = 1, 2, \ldots, n$, we get the (modified) *relative entropy* or *Kullback–Leibler divergence*

$$D_{KL}\big(P \| Q\big) = \int_X p\Big[\ln\Big(\frac{q}{p}\Big)\Big]_+ d\mu.$$

(iii) For the (convex and/or concave) functions $f_{\alpha_i}(t) = t^{\alpha_i}$, $\alpha_i \in \mathbb{R}$ for $1 \le i \le n$, the *mixed Hellinger integrals* is defined by

$$D_{(f_{\alpha_1}, f_{\alpha_2}, \ldots, f_{\alpha_n})}\big(\vec{\mathbf{P}}, \vec{\mathbf{Q}}\big) = \int_X \prod_{i=1}^n \big[p_i^{\frac{\alpha_i}{n}} q_i^{\frac{1-\alpha_i}{n}} \big] \, d\mu.$$

In particular,

$$D_{(t^\alpha, t^\alpha, \ldots, t^\alpha)}\big(\vec{\mathbf{P}}, \vec{\mathbf{Q}}\big) = \int_X \prod_{i=1}^n p_i^{\frac{\alpha}{n}} q_i^{\frac{1-\alpha}{n}} \, d\mu.$$

Those integrals are related to the Toussaint's affinity [37] and can be used to define the *mixed α-Rényi divergence*

$$D_\alpha\big(\{P_i \| Q_i\}_{i=1}^n\big) = \frac{1}{\alpha - 1} \ln\Big(\int_X \prod_{i=1}^n p_i^{\frac{\alpha}{n}} q_i^{\frac{1-\alpha}{n}} \, d\mu \Big) = \frac{1}{\alpha - 1} \ln\big[D_{(t^\alpha, t^\alpha, \ldots, t^\alpha)}\big(\vec{\mathbf{P}}, \vec{\mathbf{Q}}\big)\big].$$

The case $\alpha_i = \frac{1}{2}$, for all $i = 1, 2, \ldots, n$, gives the *mixed Bhattacharyya coefficient* or *mixed Bhattacharyya distance* of $(\vec{\mathbf{P}}, \vec{\mathbf{Q}})$,

$$D_{\big(\sqrt{t}, \sqrt{t}, \ldots, \sqrt{t}\big)}\big(\vec{\mathbf{P}}, \vec{\mathbf{Q}}\big) = \int_X \prod_{i=1}^n p_i^{\frac{1}{2n}} q_i^{\frac{1}{2n}} \, d\mu.$$

This integral is related to the Matusita's affinity [26, 27]. For more information on the corresponding f-divergences, we refer the reader to [17].

(iv) In view of existing connections between information theory and convex geometry (*e.g.*, [32, 38, 39]), we define the mixed f-divergences for convex bodies (convex and compact subsets in $\mathbb{R}^n$ with nonempty interiors) K_i with positive curvature

https://doi.org/10.4153/CMB-2016-050-3 Published online by Cambridge University Press

functions $f_{K_i}, 1 \le i \le n$, is via the measures

$$dP_{K_i} = \frac{1}{h_{K_i}^n}\, d\sigma \quad \text{and} \quad dQ_{K_i} = f_{K_i} h_{K_i}\, d\sigma, \quad 1 \le i \le n.$$

Here, σ is the spherical measure of the unit sphere S^{n-1}, $h_K(u) = \max_{x \in K}\langle x, u\rangle$ is the support function of K, and $f_K(u)$ is the curvature function of K at $u \in S^{n-1}$, the reciprocal of the Gauss curvature at x on the boundary of K with unit outer normal u. If $f_i\colon (0,\infty) \to \mathbb{R}^+, 1 \le i \le n$, are convex and/or concave functions, then

$$D_{\vec{\mathbf{f}}}\big((P_{K_1},\dots,P_{K_n}), (Q_{K_1},\dots,Q_{K_n}) \big) = \int_{S^{n-1}} \prod_{i=1}^{n} \Big[f_i\Big(\frac{1}{f_{K_i} h_{K_i}^{n+1}} \Big) f_{K_i} h_{K_i} \Big]^{\frac{1}{n}} d\sigma,$$

are the general mixed affine surface areas introduced in [41]. We refer to [33] for more details on convex bodies.

3 Inequalities

The classical Alexandrov–Fenchel inequality for mixed volumes of convex bodies is a fundamental result in (convex) geometry. A general version of this inequality for *mixed volumes* of convex bodies can be found in [1, 6, 33]. Alexandrov–Fenchel type inequalities for (mixed) affine surface areas can be found in [21, 22, 40, 41]. Now we prove an inequality for the mixed f-divergence for measures, which we call an Alexandrov–Fenchel type inequality because of its formal resemblance to be an Alexandrov–Fenchel type inequality for convex bodies.

Following [13], we say that two functions f and g are *effectively proportional* if there are constants a and b, not both zero, such that $af = bg$. Functions $f_1,\dots,f_m$ are effectively proportional if every pair $(f_i, f_j), 1 \le i, j \le m$ is effectively proportional. A null function is effectively proportional to any function. These notions will be used in the following theorems.

For a measure space (X,μ) and probability densities p_i and q_i, $1 \le i \le n$, we put

$$g_0(u) = \prod_{i=1}^{n-m} \Big[f_i\Big(\frac{p_i}{q_i}\Big) q_i \Big]^{\frac{1}{n}}, \tag{3.1}$$

and for $j = 0,\dots,m-1$,

$$g_{j+1}(u) = \Big[f_{n-j}\Big(\frac{p_{n-j}}{q_{n-j}}\Big) q_{n-j} \Big]^{\frac{1}{n}}. \tag{3.2}$$

For a vector $\vec{p}$, we let $\vec{p}^{\,n,k} = (p_1,\dots,p_{n-m}, \underbrace{p_k,\dots,p_k}_{m}), \quad k > n-m.$

Theorem 3.1 *Let (X,μ) be a measure space. For $1 \le i \le n$, let P_i and Q_i be probability measures on (X,μ) with density functions p_i and q_i, respectively μ-a.e. Let $f_i : (0,\infty) \to \mathbb{R}^+, 1 \le i \le n$, be convex functions. Then for $1 \le m \le n$,*

$$\big[D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}}) \big]^m \le \prod_{k=n-m+1}^{n} D_{\vec{f}^{\,n,k}}\big(\vec{\mathbf{P}}^{n,k}, \vec{\mathbf{Q}}^{n,k} \big).$$

Equality holds if and only if one of the functions $g_0^{\frac{1}{m}} g_i$, $1 \le i \le m$, is null or all are effectively proportional μ-a.e.

If $m = n$,

$$[D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}})]^n \le \prod_{i=1}^{n} D_{f_i}(P_i, Q_i),$$

with equality if and only if one of the functions $f_j(\frac{p_j}{q_j})q_j$, $0 \le j \le n$, is null or all are effectively proportional μ-a.e.

Remark 3.2 (i) In particular, equality holds in Theorem 3.1 if all (P_i, Q_i) coincide, and $f_i = \lambda_i f$ for some convex positive function f and $\lambda_i \ge 0$, $i = 1, 2, \dots, n$.

(ii) Theorem 3.1 still holds true if the functions f_i are concave.

Proof We let g_0 and g_{j+1}, $j = 0, \dots, m-1$ as in (3.1) and (3.2). By Hölder's inequality (see [13]),

$$\begin{aligned}
[D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}}, \vec{\mathbf{Q}})]^m &= \Big(\int_X g_0(u) g_1(u) \cdots g_m(u)\, d\mu \Big)^m \\
&= \Big(\int_X \prod_{j=0}^{m-1} \big[g_0(u) g_{j+1}(u)^m \big]^{\frac{1}{m}}\, d\mu \Big)^m \\
&\le \prod_{j=0}^{m-1} \Big(\int_X g_0(u) g_{j+1}^m(u)\, d\mu \Big) = \prod_{k=n-m+1}^{n} D_{\vec{f}^{n,k}}\big(\vec{\mathbf{P}}^{n,k}, \vec{\mathbf{Q}}^{n,k} \big).
\end{aligned}$$

Equality holds in Hölder's inequality if and only if one of the functions $g_0^{\frac{1}{m}} g_i$, $1 \le i \le m$, is null or all are effectively proportional μ-a.e. In particular, this is the case, if for all $i = 1, \dots, n$, $(P_i, Q_i) = (P, Q)$ and $f_i = \lambda_i f$ for some convex function f and $\lambda_i \ge 0$. ∎

We require some properties of f-divergences for our next result. Let $f\colon (0, \infty) \to \mathbb{R}^+$ be a convex function. By Jensen's inequality,

$$D_f(P, Q) = \int_X f\Big(\frac{p}{q} \Big) q\, d\mu \ge f\Big(\int_X p\, d\mu \Big) = f(1), \tag{3.3}$$

for all pairs of probability measures (P, Q) on (X, μ) with nonzero density functions p and q respectively μ-a.e. When f is linear, equality holds trivially in (3.3) . When f is strictly convex, equality holds true if and only if $p = q$ μ-a.e. If f is a concave function, Jensen's inequality implies

$$D_f(P, Q) = \int_X f\Big(\frac{p}{q} \Big) q\, d\mu \le f\Big(\int_X p\, d\mu \Big) = f(1) \tag{3.4}$$

for all pairs of probability measures (P, Q). Again, when f is linear, equality holds trivially. When f is strictly concave, equality holds true if and only if $p = q$ μ-a.e.

For the mixed f-divergence with concave functions, one has the following result.

Theorem 3.3 *Let (X, μ) be a measure space. For all $1 \le i \le n$, let P_i and Q_i be probability measures on X whose density functions p_i and q_i are nonzero μ-a.e. Let*

$f_i : (0,\infty) \to \mathbb{R}^+$, $1 \le i \le n$, be concave functions. Then

$$\Big[D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) \Big]^n \le \prod_{i=1}^n D_{f_i}(P_i,Q_i) \le \prod_{i=1}^n f_i(1). \tag{3.5}$$

If in addition, all f_i are strictly concave, equality holds if and only if there is a probability density p such that for all $i = 1,2,\dots,n$, $p_i = q_i = p$, $\mu - a.e.$

Proof Theorem 3.1 and Remark 3.2 imply that for all concave functions f_i,

$$[D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}})]^n \le \prod_{i=1}^n D_{f_i}(P_i,Q_i) \le \prod_{i=1}^n f_i(1),$$

where the second inequality follows from inequality (3.4) and $f_i \ge 0$.

Suppose now that for all i, $p_i = q_i = p$, μ-a.e., where p is a fixed probability density. Then equality holds trivially in (3.5). Conversely, suppose that equality holds in (3.5). Then, in particular, equality holds in Jensen's inequality, which, as noted above, happens if and only if $p_i = q_i$ for all i. Thus,

$$D_{\vec{\mathbf{f}}}(\vec{\mathbf{P}},\vec{\mathbf{Q}}) = \Big(\prod_{i=1}^n [f_i(1)]^{1/n} \Big) \int_X q_1^{1/n} \cdots q_n^{1/n}\, d\mu.$$

Note also that if all $f_i\colon (0,\infty) \to \mathbb{R}^+$ are strictly concave, then $f_i(1) \ne 0$ for all $1 \le i \le n$. Equality characterization in Hölder's inequality implies that all q_i are effectively proportional μ-a.e. As all q_i are probability measures, they are all equal (μ-a.e.) to a probability measure with density function (say) p. ∎

Remark If $f_i(t) = a_i t + b_i$ are all linear and positive, then equality holds if and only if all p_i, q_i are equal (μ-a.e.) as convex combinations, *i.e.*, if and only if for all i, j,

$$\frac{a_i}{a_i+b_i} p_i + \frac{b_i}{a_i+b_i} q_i = \frac{a_j}{a_j+b_j} p_j + \frac{b_j}{a_j+b_j} q_j, \quad \mu - \text{a.e.}$$

4 The i-th Mixed f-divergence

Let (X,μ) be a measure space. Throughout this section, we assume that the functions

$$f_1, f_2\colon (0,\infty) \longrightarrow \{x \in \mathbb{R} : x > 0\},$$

are convex or concave and that P_1, P_2, Q_1, Q_2 are probability measures on X with density functions p_1, p_2, q_1, q_2 that are nonzero μ-a.e. We also write

$$\vec{f} = (f_1, f_2), \quad \vec{P} = (P_1,P_2), \quad \vec{Q} = (Q_1,Q_2).$$

Definition 4.1 Let $i \in \mathbb{R}$. The i-th mixed f-divergence for $(\vec{f},\vec{P},\vec{Q})$, denoted by $D_{\vec{f}}(\vec{P},\vec{Q};i)$, is defined as

$$D_{\vec{f}}(\vec{P},\vec{Q};i) = \int_X \Big[f_1\Big(\frac{p_1}{q_1}\Big) q_1 \Big]^{\frac{i}{n}} \Big[f_2\Big(\frac{p_2}{q_2}\Big) q_2 \Big]^{\frac{n-i}{n}} d\mu. \tag{4.1}$$

Remarks Note that the i-th mixed f-divergence is defined for any combination of convexity and concavity of f_1 and f_2, namely, both f_1 and f_2 concave, or both f_1 and f_2 convex, or one is convex the other is concave.

It is easily checked that

$$D_{\vec{f}}(\vec{P},\vec{Q};i) = D_{(f_2,f_1)}\big((P_2,Q_2),(P_1,Q_1);n-i\big).$$

If $0 \le i \le n$ is an integer, then the triple (f_1,P_1,Q_1) appears i-times, while the triple (f_2,P_2,Q_2) appears $(n-i)$ times in $D_{\vec{f}}(\vec{P},\vec{Q};i)$. Note that if $i=0$, then $D_{\vec{f}}(\vec{P},\vec{Q};i) = D_{f_2}(P_2,Q_2)$, and if $i=n$, then $D_{\vec{f}}(\vec{P},\vec{Q};i) = D_{f_1}(P_1,Q_1)$.

Another special case is when $P_2 = Q_2 = \mu$ almost everywhere and μ is also a probability measure. Then such an i-th mixed f-divergence, denoted by $D\big((f_1,P_1,Q_1),i;f_2\big)$, has the form

$$D\big((f_1,P_1,Q_1),i;f_2\big) = [f_2(1)]^{1-i/n}\int_X\Big[f_1\Big(\frac{p_1}{q_1}\Big)q_1\Big]^{\frac{i}{n}}\,d\mu.$$

Examples and Applications (i) For $f(t) = |t-1|$, we get the *i-th mixed total variation*

$$D_{TV}\big(\vec{P},\vec{Q};i\big) = \int_X |p_1-q_1|^{\frac{i}{n}}|p_2-q_2|^{\frac{n-i}{n}}\,d\mu.$$

(ii) For $f_1(t) = f_2(t) = [t\ln t]_+$, we get the (modified) *i-th mixed relative entropy* or *i-th mixed Kullback–Leibler divergence*

$$D_{KL}\big(\vec{P},\vec{Q};i\big) = \int_X\Big[p_1\ln\Big(\frac{p_1}{q_1}\Big)\Big]_+^{\frac{i}{n}}\Big[p_2\ln\Big(\frac{p_2}{q_2}\Big)\Big]_+^{\frac{n-i}{n}}\,d\mu.$$

(iii) For the convex or concave functions $f_{\alpha_j}(t) = t^{\alpha_j}$, $j=1,2$, we get the *i-th mixed Hellinger integrals*

$$D_{(f_{\alpha_1},f_{\alpha_2})}\big(\vec{P},\vec{Q};i\big) = \int_X (p_1^{\alpha_1}q_1^{1-\alpha_1})^{\frac{i}{n}}(p_2^{\alpha_2}q_2^{1-\alpha_2})^{\frac{n-i}{n}}\,d\mu.$$

In particular, for $\alpha_j = \alpha$, for $j=1,2$,

$$D_{(f_\alpha,f_\alpha)}\big(\vec{P},\vec{Q};i\big) = \int_X (p_1^{\alpha}q_1^{1-\alpha})^{\frac{i}{n}}(p_2^{\alpha}q_2^{1-\alpha})^{\frac{n-i}{n}}\,d\mu.$$

This integral can be used to define the *i-th mixed α-Rényi divergence*

$$D_\alpha\big(\vec{P},\vec{Q};i\big) = \frac{1}{\alpha-1}\ln\big[D_{(f_\alpha,f_\alpha)}\big(\vec{P},\vec{Q};i\big)\big].$$

The case $\alpha_i = \frac{1}{2}$ for all i gives

$$D_{(\sqrt{t},\sqrt{t})}\big(\vec{P},\vec{Q};i\big) = \int_X (p_1q_1)^{\frac{i}{2n}}(p_2q_2)^{\frac{n-i}{2n}}\,d\mu,$$

the *i-th mixed Bhattacharyya coefficient* or *i-th mixed Bhattacharyya distance* of p_i and q_i.

(iv) Important applications are again in the theory of convex bodies. As in Section 2, let K_1 and K_2 be convex bodies with positive curvature function. For $l=1,2$, let

$$dP_{K_l} = \frac{1}{h_{K_l}^n}\,d\sigma \quad\text{and}\quad dQ_{K_l} = f_{K_l}h_{K_l}\,d\sigma.$$

https://doi.org/10.4153/CMB-2016-050-3 Published online by Cambridge University Press

Let $f_l\colon (0,\infty) \to \mathbb{R}$, $l = 1,2$, be positive convex functions. Then we define the *i-th mixed f-divergence* for convex bodies K_1 and K_2 by

$$D_{\vec{f}}\big((P_{K_1}, P_{K_2}), (Q_{K_1}, Q_{K_2}); i\big) = \int_{S^{n-1}} \Big[f_1\Big(\frac{1}{f_{K_1} h_{K_1}^{n+1}}\Big) f_{K_1} h_{K_1}\Big]^{\frac{i}{n}} \Big[f_2\Big(\frac{1}{f_{K_2} h_{K_2}^{n+1}}\Big) f_{K_2} h_{K_2}\Big]^{\frac{n-i}{n}} \, d\sigma.$$

These are the general i-th mixed affine surface areas introduced in [41].

The following result holds for all possible combinations of convexity and concavity of f_1 and f_2.

Proposition 4.2 *Let $\vec{f}, \vec{P}, \vec{Q}$ be as above. If $j \le i \le k$ or $k \le i \le j$, then*

$$D_{\vec{f}}(\vec{P}, \vec{Q}; i) \le \big[D_{\vec{f}}(\vec{P}, \vec{Q}; j)\big]^{\frac{k-i}{k-j}} \times \big[D_{\vec{f}}(\vec{P}, \vec{Q}; k)\big]^{\frac{i-j}{k-j}}.$$

Equality holds trivially if $i = k$ or $i = j$. Otherwise, equality holds if and only if one of the functions $f_i(p_i/q_i)q_i$, $i = 1,2$, is null, or $f_1(p_1/q_1)q_1$ and $f_2(p_2/q_2)q_2$ are effectively proportional μ-a.e. In particular, this holds if $(P_1, Q_1) = (P_2, Q_2)$ and $f_1 = \lambda f_2$ for some $\lambda > 0$.

Proof By formula (4.1), one has

$$\begin{aligned} D_{\vec{f}}(\vec{P}, \vec{Q}; i) &= \int_X \Big[f_1\Big(\frac{p_1}{q_1}\Big) q_1\Big]^{\frac{i}{n}} \Big[f_2\Big(\frac{p_2}{q_2}\Big) q_2\Big]^{\frac{n-i}{n}} d\mu \\ &= \int_X \Big\{ \Big[f_1\Big(\frac{p_1}{q_1}\Big) q_1\Big]^{\frac{j}{n}} \Big[f_2\Big(\frac{p_2}{q_2}\Big) q_2\Big]^{\frac{n-j}{n}} \Big\}^{\frac{k-i}{k-j}} \\ &\qquad \times \Big\{ \Big[f_1\Big(\frac{p_1}{q_1}\Big) q_1\Big]^{\frac{k}{n}} \Big[f_2\Big(\frac{p_2}{q_2}\Big) q_2\Big]^{\frac{n-k}{n}} \Big\}^{\frac{i-j}{k-j}} d\mu \\ &\le \big[D_{\vec{f}}(\vec{P}, \vec{Q}; j)\big]^{\frac{k-i}{k-j}} \times \big[D_{\vec{f}}(\vec{P}, \vec{Q}; k)\big]^{\frac{i-j}{k-j}}, \end{aligned}$$

where the last inequality follows from Hölder's inequality and formula (4.1). The equality characterization follows from the one in Hölder inequality. In particular, if $(P_1, Q_1) = (P_2, Q_2)$, and $f_1 = \lambda f_2$ for some $\lambda > 0$, equality holds. ∎

Corollary 4.3 *Let f_1 and f_2 be positive, concave functions on $(0,\infty)$. Then for all $\vec{P}, \vec{Q}$ and for all $0 \le i \le n$,*

$$\big[D_{\vec{f}}(\vec{P}, \vec{Q}; i)\big]^n \le [f_1(1)]^i [f_2(1)]^{n-i}.$$

If in addition, f_1 and f_2 are strictly concave, equality holds if and only if $p_1 = p_2 = q_1 = q_2$ μ-a.e.

Proof Let $j = 0$ and $k = n$ in Proposition 4.2. Then for all $0 \le i \le n$,

$$\big[D_{\vec{f}}(\vec{P}, \vec{Q}; i)\big]^n \le \big[D_{f_1}(P_1, Q_1)\big]^i \big[D_{f_2}(P_2, Q_2)\big]^{n-i} \le \big[f_1(1)\big]^i \big[f_2(1)\big]^{n-i},$$

where the last inequality follows from inequality (3.4).

To have equality, the above inequalities should be equalities. Proposition 4.2 implies that $f_1(p_1/q_1)q_1$ and $f_2(p_2/q_2)q_2$ are effectively proportional μ-a.e. As both f_1 and f_2 are strictly concave, Jensen's inequality requires that $p_1 = q_1$ and $p_2 = q_2$ μ-a.e. Therefore, equality holds if and only if $f_1(1)q_1$ and $f_2(1)q_2$ are effectively proportional μ-a.e. As both $f_1(1)$ and $f_2(1)$ are not zero, equality holds if and only if $p_1 = p_2 = q_1 = q_2$ μ-a.e. ∎

Remark If $f_1(t) = a_1t + b_1$ and $f_2(t) = a_2t + b_2$ are both linear, equality holds in Corollary 4.3 if and only if p_i, q_i, $i = 1, 2$, are equal as convex combinations, *i.e.*,

$$\frac{a_1}{a_1+b_1}p_1 + \frac{b_1}{a_1+b_1}q_1 = \frac{a_2}{a_2+b_2}p_2 + \frac{b_2}{a_2+b_2}q_2, \quad \mu - \text{a.e.}$$

This proof can be used to establish the following result for $D((f_1, P_1, Q_1), i; f_2)$.

Corollary 4.4 *Let (X, μ) be a probability space. Let f_1 be a positive concave function on $(0, \infty)$. Then for all P_1, Q_1, for all (concave or convex) positive functions f_2, and for all $0 \le i \le n$,*

$$\big[D\big((f_1, P_1, Q_1), i; f_2\big)\big]^n \le [f_1(1)]^i[f_2(1)]^{n-i}.$$

If f_1 is strictly concave, equality holds if and only if $P_1 = Q_1 = \mu$. When $f_1(t) = at + b$ is linear, equality holds if and only if $ap_1 + bq_1 = a + b$ μ-a.e.

Corollary 4.5 *Let f_1 be a positive convex function and let f_2 be a positive concave function on $(0, \infty)$. Then, for all $\vec{P}, \vec{Q}$, and for all $k \ge n$,*

$$\big[D_{\vec{f}}\big(\vec{P}, \vec{Q}; k\big)\big]^n \ge [f_1(1)]^k[f_2(1)]^{n-k}.$$

If in addition, f_1 is strictly convex and f_2 is strictly concave, equality holds if and only if $p_1 = p_2 = q_1 = q_2$ μ-a.e.

Proof On the right-hand side of Proposition 4.2, let $i = n$ and $j = 0$. Let $k \ge n$. Then

$$\big[D_{\vec{f}}\big(\vec{P}, \vec{Q}; k\big)\big]^n \ge \big[D_{f_1}(P_1, Q_1)\big]^k\big[D_{f_2}(P_2, Q_2)\big]^{n-k} \ge \big[f_1(1)\big]^k\big[f_2(1)\big]^{n-k}.$$

Here, the last inequality follows from inequalities (3.3), (3.4) and $k \ge n$. To have equality, the above inequalities should be equalities. Proposition 4.2 implies that $f_1(p_1/q_1)q_1$ and $f_2(p_2/q_2)q_2$ are effectively proportional μ-a.e. As f_1 is strictly convex and f_2 is strictly concave, Jensen's inequality implies that $p_1 = q_1$ and $p_2 = q_2$ μ-a.e. Therefore, as both $f_1(1)$ and $f_2(1)$ are not zero, equality holds if and only if $p_1 = p_2 = q_1 = q_2$ μ-a.e. ∎

Remark If $f_1(t) = a_1t + b_1$ and $f_2(t) = a_2t + b_2$ are both linear, equality holds in Corollary 4.5 if and only if p_i, q_i, $i = 1, 2$, are equal μ-a.e. as convex combinations, *i.e.*,

$$\frac{a_1}{a_1+b_1}p_1 + \frac{b_1}{a_1+b_1}q_1 = \frac{a_2}{a_2+b_2}p_2 + \frac{b_2}{a_2+b_2}q_2, \quad \mu - \text{a.e.}$$

This proof can be used to establish the following result for $D((f_1, P_1, Q_1), k; f_2)$.

Corollary 4.6 *Let (X,μ) be a probability space. Let f_1 be a positive convex function on $(0,\infty)$. Then for all P_1, Q_1, for all (positive concave or convex) functions f_2, and for all $k \geq n$,*

$$\Big[D\big((f_1,P_1,Q_1),k;f_2\big)\Big]^n \geq [f_1(1)]^k [f_2(1)]^{n-k}.$$

If f_1 is strictly convex, equality holds if and only if $P_1 = Q_1 = \mu$. When $f_1(t) = at+b$ is linear, equality holds if and only if $ap_1 + bq_1 = a+b$ μ-a.e.

Corollary 4.7 *Let f_1 be a positive concave function and let f_2 be a positive convex function on $(0,\infty)$. Then for all $\vec{P}, \vec{Q}$, and for all $k \leq 0$,*

$$\Big[D_{\vec{f}}(\vec{P},\vec{Q};k)\Big]^n \geq [f_1(1)]^k [f_2(1)]^{n-k}.$$

If in addition, f_1 is strictly concave and f_2 is strictly convex, equality holds if and only if $p_1 = p_2 = q_1 = q_2$ μ-a.e.

Proof Let $i=0$ and $j=n$ in Proposition 4.2. Then

$$\Big[D_{\vec{f}}\big(\vec{P},\vec{Q};k\big)\Big]^n \geq \Big[D_{f_1}(P_1,Q_1)\Big]^k \Big[D_{f_2}(P_2,Q_2)\Big]^{n-k} \geq \Big[f_1(1)\Big]^k \Big[f_2(1)\Big]^{n-k}.$$

Here, the last inequality follows from inequalities (3.3), (3.4), and $k \leq 0$.

To have equality, the above inequalities should be equalities. Proposition 4.2 implies that $f_1(p_1/q_1)q_1$ and $f_2(p_2/q_2)q_2$ are effectively proportional μ-a.e. As f_1 is strictly concave and f_2 is strictly convex, Jensen's inequality requires that $p_1 = q_1$ and $p_2 = q_2$. Therefore, equality holds if and only if $f_1(1)q_1$ and $f_2(1)q_2$ are effectively proportional μ-a.e. As both $f_1(1)$ and $f_2(1)$ are not zero, equality holds if and only if $p_1 = p_2 = q_1 = q_2$ μ-a.e. ∎

This proof can be used to establish the following result for $D\big((f_1,P_1,Q_1),k;f_2\big)$.

Corollary 4.8 *Let f_1 be a concave function on $(0,\infty)$. Then for all P_1, Q_1, for all (concave or convex) functions f_2, and for all $k \leq 0$,*

$$\Big[D\big((f_1,P_1,Q_1),k;f_2\big)\Big]^n \geq \Big[f_1(1)\Big]^k \Big[f_2(1)\Big]^{n-k}.$$

If f_1 is strictly concave, equality holds if and only if $P_1 = Q_1 = \mu$. When $f_1(t) = at+b$ is linear, equality holds if and only if $ap_1 + bq_1 = a+b$ μ-a.e.

References

[1] A. D. Aleksandrov, *On the theory of mixed volumes of convex bodies. II. New inequalities between mixed volumes and their applications.* (Russian) Mat. Sb. (N. S.) **2**(1937), 1205–1238.

[2] S. M. Ali and D. Silvey, *A general class of coefficients of divergence of one distribution from another.* J. Roy. Statist. Soc. Ser. B **28**(1966), 131–142.

[3] A. R. Barron, L. Györfi, and E. C. van der Meulen, *Distribution estimates consistent in total variation and two types of information divergence.* IEEE Trans. Inform. Theory **38**(1990), 1437–1454.

[4] M. Basseville, *Divergence measures for statistical data processing.* Technical Report PI 1961, IRISA, November 2010. http://hal.inria.fr/inria-00542337/fr/

[5] A. Bhattacharyya, *On some analogues to the amount of information and their uses in statistical estimation.* Sankhya **8**(1946), 1–14.

[6] H. Busemann, *Convex surfaces.* Interscience Tracts in Pure and Applied Mathematics, 6, Interscience, London, 1958.

[7] T. Cover and J. Thomas, *Elements of information theory.* second ed., Wiley-Interscience, (John Wiley and Sons), Hoboken, NJ, 2006.
[8] I. Csiszár, *Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten.* Magyar Tud. Akad. Mat. Kutató Int. Közl. **8**(1963), 85–108.
[9] R. J. Gardner, *The Brunn-Minkowski Inequality*. Bull. Amer. Math. Soc. **39**(2002), no. 3, 355–405 . http://dx.doi.org/10.1090/S0273-0979-02-00941-2
[10] O. G. Guleryuz, E. Lutwak, D. Yang, and G. Zhang, *Information theoretic inequalities for contoured probability distributions*. IEEE Trans. Inform. Theory **48**(2002), 2377–2383. http://dx.doi.org/10.1109/TIT.2002.800496
[11] L. Györfi and T. Nemetz, *f-dissimilarity: A general class of separation measures of several probability measures.* In: Topics in information theory, Colloquia Mathematica Societatis János Bolyai, 16, North-Holland, 1975, pp. 309–321.
[12] ———, *f-dissimilarity: A generalization of the affinity of several distributions.* Ann. Inst. Statist. Math. **30**(1978), 105–113. http://dx.doi.org/10.1007/BF02480206
[13] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*. second ed., Cambridge University Press, Cambridge, 1952.
[14] P. Harremoes and F. Topsoe, *Inequalities between entropy and the index of coincidence derived from information diagrams*. IEEE Trans. Inform. Theory **47**(2001), no. 7, 2944–2960. http://dx.doi.org/10.1109/18.959272
[15] J. Jenkinson and E. Werner, *Relative entropies for convex bodies*. Trans. Amer. Math. Soc. **366**(2014), no. 6, 2889–2906. http://dx.doi.org/10.1090/S0002-9947-2014-05788-7
[16] S. Kullback and R. A. Leibler, *On information and sufficiency.* Ann. Math. Statistics **22**(1951), 79–86. http://dx.doi.org/10.1214/aoms/1177729694
[17] F. Liese and I. Vajda, *On divergences and information in statistics and information theory.* IEEE Trans. Inform. Theory **52**(2006), 4394–4412. http://dx.doi.org/10.1109/TIT.2006.881731
[18] M. Ludwig, *General affine surface areas*. Adv. Math. **224**(2010), no. 6, 2346–2360. http://dx.doi.org/10.1016/j.aim.2010.02.004
[19] M. Ludwig and M. Reitzner, *A characterization of affine surface area*. Adv. Math. **147**(1999), no. 1, 138–172. http://dx.doi.org/10.1006/aima.1999.1832
[20] ———, *A classification of $SL(n)$ invariant valuations*. Annals of Math. **172**(2010), no. 2, 1223–1271. http://dx.doi.org/10.4007/annals.2010.172.1223
[21] E. Lutwak, *Mixed affine surface area.* J. Math. Anal. Appl. **125**(1987), no. 2, 351–360. http://dx.doi.org/10.1016/0022-247X(87)90097-7
[22] E. Lutwak, *The Brunn-Minkowski-Firey theory. II. affine and geominimal surface areas.* Adv. Math. **118**(1996), no. 2, 244–294. http://dx.doi.org/10.1006/aima.1996.0022
[23] E. Lutwak, D. Yang, and G. Zhang, *The Cramer-Rao inequality for star bodies*. Duke Math. J. **112**(2002), 59–81. http://dx.doi.org/10.1215/S0012-9074-02-11212-5
[24] ———, *Moment-entropy inequalities*. Ann. Probab. **32**(2004), 757–774. http://dx.doi.org/10.1214/aop/1079021463
[25] ———, *Cramer-Rao and moment-entropy inequalities for Renyi entropy and generalized Fisher information*. IEEE Trans. Inform. Theory **51**(2005), 473–478. http://dx.doi.org/10.1109/TIT.2004.840871
[26] K. Matusita, *On the notion of affinity of several distributions and some of its applications.* Ann. Inst. Statist. Math. **19**(1967), 181–192. http://dx.doi.org/10.1007/BF02911675
[27] ———, *Some properties of affinity and applications.* Ann. Inst. Statist. Math. **23**(1971), 137–155. http://dx.doi.org/10.1007/BF02479219
[28] M. L. Menéndez, J. A. Pardo, L. Pardo, and K. Zografos, *A preliminary test in classification and probabilities of misclassification.* Statistics **39**(2005), 183–205. http://dx.doi.org/10.1080/02331880500097986
[29] D. Morales, L. Pardo, and K. Zografos, *Informational distances and related statistics in mixed continuous and categorical variables.* J. Statist. Plann. Inference **75**(1998), 47–63. http://dx.doi.org/10.1016/S0378-3758(98)00120-7
[30] T. Morimoto, *Markov processes and the H-theorem.* J. Phys. Soc. Jap. **18**(1963), 328–331. http://dx.doi.org/10.1143/JPSJ.18.328
[31] 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), 639–653. http://dx.doi.org/10.1007/BF02517812
[32] G. Paouris and E. Werner, *Relative entropy of cone measures and L_p centroid bodies.* Proc. London Math. Soc. **104**(2012), 253–286. http://dx.doi.org/10.1112/plms/pdr030

[33] R. Schneider, *Convex bodies: the Brunn–Minkowski theory.* Cambridge University Press, Cambridge, 1993. http://dx.doi.org/10.1017/CBO9780511526282
[34] C. Schütt and E. Werner, *Surface bodies and p-affine surface area.* Adv. Math. **187**(2004), no. 1, 98–145. http://dx.doi.org/10.1016/j.aim.2003.07.018
[35] A. Sgarro, *Informational divergence and the dissimilarity of probability distributions.* Calcolo **18**(1981), 293–302. http://dx.doi.org/10.1007/BF02576360
[36] R. Sibson, *Information radius.* Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **14**(1969), 149–160. http://dx.doi.org/10.1007/BF00537520
[37] G. T. Toussaint, *Some properties of Matusita's measure of affinity of several distributions.* Ann. Inst. Statist. Math. **26**(1974), 389–394. http://dx.doi.org/10.1007/BF02479834
[38] E. Werner, *Rényi Divergence and L_p-affine surface area for convex bodies.* Adv. Math. **230**(2012), 1040–1059. http://dx.doi.org/10.1016/j.aim.2012.03.015
[39] ———, *f-Divergence for convex bodies.* Proceedings of the "Asymptotic Geometric Analysis" workshop, Fields Inst. Commun., Springer, New York, 2012, pp. 381–395. http://dx.doi.org/10.1007/978-1-4614-6406-8_18
[40] E. Werner and D. Ye, *Inequalities for mixed p-affine surface area.* Math. Ann. **347**(2010), 703–737. http://dx.doi.org/10.1007/s00208-009-0453-2
[41] D. Ye, *Inequalities for general mixed affine surface areas.* J. London Math. Soc. **85**(2012), 101–120. http://dx.doi.org/10.1112/jlms/jdr043
[42] K. Zografos, *f-dissimilarity of several distributions in testing statistical hypotheses.* Ann. Inst. Statist. Math. **50**(1998), 295–310. http://dx.doi.org/10.1023/A:1003443215838

Department of Mathematics, Case Western Reserve University, Cleveland, Ohio 44106, USA
e-mail: elisabeth.werner@case.edu

Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, Newfoundland A1C 5S7
e-mail: deping.ye@mun.ca