KYBERNETIKA — VOLUME 32 (1996), NUMBER 4, PAGES 389–393

ON A CLASS OF PERIMETER–TYPE DISTANCES OF PROBABILITY DISTRIBUTIONS

FERDINAND ÖSTERREICHER

The class I_{f_p}, $p \in (1,\infty]$, of f-divergences investigated in this paper generalizes an f-divergence introduced by the author in [9] and applied there and by Reschenhofer and Bomze [11] in different areas of hypotheses testing. The main result of the present paper ensures that, for every $p \in (1,\infty)$, the square root of the corresponding divergence defines a distance on the set of probability distributions. Thus it generalizes the respecting statement for $p = 2$ made in connection with Example 4 by Kafka, Österreicher and Vincze in [6].

From the former literature on the subject the maximal powers of f-divergences defining a distance are known for the subsequent classes. For the class of Hellinger-divergences given in terms of $f^{(s)}(u) = 1+u-(u^s+u^{1-s})$, $s \in (0,1)$, already Csiszár and Fischer [3] have shown that the maximal power is $\min(s, 1-s)$. For the following two classes the maximal power coincides with their parameter. The class given in terms of $f_{(\alpha)}(u) = |1-u^\alpha|^{\frac{1}{\alpha}}$, $\alpha \in (0,1]$, was investigated by Boekee [2]. The previous class and this one have the special case $s = \alpha = \frac{1}{2}$ in common. This famous case is attributed to Matusita [8]. The class given by $\varphi_\alpha(u) = |1-u|^{\frac{1}{\alpha}}(1+u)^{1-\frac{1}{\alpha}}$, $\alpha \in (0,1]$, and investigated in [6], Example 3, contains the wellknown special case $\alpha = \frac{1}{2}$ introduced by Vincze [13].

1. INTRODUCTION

Let $(\Omega, \mathcal{A})$ be a nondegenerate measurable space (i.e. $|\mathcal{A}| > 2$ and hence $|\Omega| > 1$) and let $\mathcal{M}_1(\Omega, \mathcal{A})$ be the set of probability distributions on $(\Omega, \mathcal{A})$. Furthermore let $\mathcal{F}$ be the set of convex functions $f : \mathbb{R}_+ \to \mathbb{R}$ which are continuous at 0. And let the function $f^* \in \mathcal{F}$ be defined by

$$f^*(u) = u \cdot f\left(\frac{1}{u}\right) \quad \text{for } u \in (0,\infty).$$

Remark 1. Owing to the continuity of f and f^* at 0 and by setting $0 \cdot f\left(\frac{0}{0}\right) = 0$ for all $f \in \mathcal{F}$ it holds

$$x \cdot f^*\left(\frac{y}{x}\right) = y \cdot f\left(\frac{x}{y}\right) \quad \text{for all } x, y \in \mathbb{R}_+ .$$

Definition (cf. Csiszár [4] and Ali and Silvey [1]). Let $Q_0, Q_1 \in \mathcal{M}_1(\Omega, \mathcal{A})$. Then

$$I_f(Q_1, Q_0) = \int f\left(\frac{q_1}{q_0}\right) \cdot q_0 \, \mathrm{d}\mu$$

is called f–divergence of Q_0 and Q_1. (As usual, q_1 and q_0 denote the Radon–Nikodym–derivatives of Q_1 and Q_0 with respect to a dominating σ–finite measure μ.)

In the sequel we briefly restate those results from [6] which are basic for the statement and proof of the main result of this paper. For further informations on f–divergences we refer to the monograph [7] by Liese and Vajda and the paper [12] of Vajda and Österreicher.

Provided

(f1) $f(1) = 0$ and f is strictly convex at 1 and

(f2) $f^*(u) \equiv f(u)$

it holds

(M1) $I_f(Q_1, Q_0) \geq 0$ with equality iff $Q_0 = Q_1$ $\forall\, Q_0, Q_1 \in \mathcal{M}_1(\Omega, \mathcal{A})$,

(M2) $I_f(Q_1, Q_0) = I_f(Q_0, Q_1)$ $\forall\, Q_0, Q_1 \in \mathcal{M}_1(\Omega, \mathcal{A})$

respectively. If, in addition to (f1) and (f2), there exists an $\alpha \in (0, 1]$, such that

(f3,α) the function $h(u) = \dfrac{(1-u^\alpha)^{\frac{1}{\alpha}}}{f(u)}$, $u \in [0, 1)$,

is (not neccessarily strictly) decreasing,

then, according to [6], Theorems 1 and 2, the power

$$\rho_\alpha(Q_0, Q_1) = [I_f(Q_1, Q_0)]^\alpha$$

of the f–divergence satisfies the triangle inequality

(M3) $\rho_\alpha(Q_0, Q_1) \leq \rho_\alpha(Q_0, Q_2) + \rho_\alpha(Q_2, Q_1)$ $\forall\, Q_0, Q_1, Q_2 \in \mathcal{M}_1(\Omega, \mathcal{A})$.

Remark 2. Note that by virtue of Jensen's inequality

$$\frac{f(u) + f^*(u)}{1+u} = \frac{1}{1+u} \cdot f(u) + \frac{u}{1+u} \cdot f\left(\frac{1}{u}\right) \geq f(1).$$

Therefore (f1) and (f2) imply $f(u) > 0$ for all $u \in \mathbb{R}_+ \setminus \{1\}$ and hence $f(0) \in (0, \infty)$. Moreover, it can be easily seen that, provided (f3,β) is satisfied for $\beta = \alpha \in (0, 1]$, it is also satisfied for every $\beta \in (0, 1]$.

The following Remark is a consequence of [6], Propositions 5 and 6.

Remark 3. Let (f1) and (f2) hold true and let $\alpha_0 \in (0,1]$ be the maximal α for which (f3,α) is satisfied. Then the following statement concerning α_0 can be made. Let $k_0,\ k_1,\ c_0,\ c_1 \in (0,\infty)$ be such that

$$\begin{aligned} f(0)\cdot(1+u)-f(u) &\sim c_0\cdot u^{k_0} && \text{for } u\downarrow 0 \quad \text{and} \\ f(u) &\sim c_1\cdot|u-1|^{k_1} && \text{for } u\uparrow 1 \end{aligned}$$

then $k_0 \le 1,\ k_1 \ge 1$ and $\alpha_0 \le \min\left(k_0, \frac{1}{k_1}\right) \le 1$.

2. THE MAIN RESULT

First we are going to show that f–divergences can be defined in terms of the following class of functions

$$f_p(u) = \begin{cases} (1+u^p)^{\frac{1}{p}} - 2^{\frac{1}{p}-1}\cdot(1+u) & \text{for } \ p\in(1,\infty) \\ \dfrac{|u-1|}{2} & \text{for } \ p=\infty \end{cases}, \quad u\in\mathbb{R}_+$$

which satisfies $\lim_{p\to\infty} f_p(u) = f_\infty(u)$.

Lemma 1. $f_p \in \mathcal{F}$ and satisfies (f1) and (f2) for all $p\in(1,\infty]$.

Proof. Since this assertion is obvious for the case $p=\infty$, let us assume $p\in(1,\infty)$ from now on. For this case

$$\lim_{u\downarrow 0} f_p(u) = f_p(0) = 1-2^{\frac{1}{p}-1}\in(0,\infty)\,, \quad f_p(1)=0\,,$$

(1) $$f_p'(u) = (1+u^p)^{\frac{1}{p}-1}\cdot u^{p-1} - 2^{\frac{1}{p}-1} \quad \text{and hence}$$
$$f_p'(1) = 0 \quad \text{and}$$

(2) $$f_p''(u) = (p-1)\cdot(1+u^p)^{\frac{1}{p}-2}\cdot u^{p-2} > 0 \quad \forall\, u\in(0,\infty) \quad \text{and hence}$$
$$f_p''(1) = (p-1)\cdot 2^{\frac{1}{p}-2}\,.$$

Therefore f_p is an element of $\mathcal{F}$ satisfying (f1). The validity of $f_p^*(u)\equiv f_p(u)$ is obvious. □

Remark 3 provides an upper bound for the subset of those $\alpha\in(0,1]$, for which (f3,α) may hold.

Remark 4. Owing to

$$\begin{aligned} f_p(0)\cdot(1+u)-f_p(u) &= 1+u-(1+u^p)^{\frac{1}{p}} \sim u && \text{for } u\downarrow 0\,, \\ f_p(u) &\sim (p-1)\cdot 2^{\frac{1}{p}-3}\cdot(u-1)^2 && \text{for } u\uparrow 1\,, \end{aligned}$$

(the latter being a consequence of (1) and (2)), the maximal $\alpha\in(0,1]$ satisfying (f3,α) – if there is any – must be $\alpha_0 \le \frac{1}{2}$.

Interpretation of the f–divergences under consideration. Let

$$R(Q_0, Q_1) = co\{(Q_0(A), Q_1(A^c)), A \in \mathcal{A}\}$$

be the risk set of the testing problem $(Q_0, Q_1) \in \mathcal{M}_1(\Omega, \mathcal{A})^2$ (whereby "*co*" means "the convex hull of"). Then the corresponding f–divergence

$$I_{f_p}(Q_1, Q_0) = \begin{cases} \int ({q_1}^p + {q_0}^p)^{\frac{1}{p}}\, \mathrm{d}\mu - 2^{\frac{1}{p}} & \text{for} \quad p \in (1, \infty) \\ \frac{1}{2} \int |q_1 - q_0|\, \mathrm{d}\mu & \text{for} \quad p = \infty \end{cases}$$

can be interpreted as the difference or the arc lengths of the lower boundary of the risk set and the diagonal

$$D = \{(x, y) \in [0, 1]^2 : x + y = 1\},$$

both measured in terms of the l_p–norm in $\mathbb{R}^2$. We denote the arc length in question by l_p–arc length since it coincides for $p = 2$ with the ordinary arc length. For further reading on the geometric point of view we refer to Feldman and Österreicher [5] and the entry [10] of the author.

For the limiting case $p = \infty$ the corresponding f–divergence $I_{f_\infty}(Q_1, Q_0)$ is half of the well-known variation distance. For $p = 2$ it has been shown in [6] that the square root of the corresponding f–divergence $I_{f_2}(Q_1, Q_0)$ is also a distance. In the sequel we are going to show the following generalization of the latter which may be conjectured from Remark 4.

Theorem. For every $p \in (1, \infty)$ the square root of the f–divergence $I_{f_p}(Q_1, Q_0)$ defines a distance on $\mathcal{M}_1(\Omega, \mathcal{A})$.

By virtue of Lemma 1 and [6], Theorems 1 and 2, the proof is reduced to that of the following Lemma.

Lemma 2. Let $p \in (1, \infty)$. Then the function

$$h_p(u) = \frac{(\sqrt{u} - 1)^2}{f_p(u)}, \quad u \in [0, 1),$$

is (strictly) decreasing.

Proof. Because of

$$h_p'(u) = \left(\frac{1}{\sqrt{u}} - 1\right) \cdot \frac{1}{f_p^2(u)} \cdot \phi_p(u)$$

with

$$\begin{aligned} \phi_p(u) &= -\left[f_p(u) + (\sqrt{u} - u) \cdot f_p'(u)\right] \\ &= 2^{\frac{1}{p}-1}(1 + u^{\frac{1}{2}}) - (1 + u^p)^{\frac{1}{p}-1} \cdot (1 + u^{p-\frac{1}{2}}) \end{aligned}$$

it suffices to show $\phi_p(u) < 0$ for all $u \in (0,1)$. Owing to $\phi_p(1) = 0$ it suffices to show that the functions ψ_p defined by $\psi_p(u) = \sqrt{u} \cdot \phi_p'(u)$ satisfy

$$\psi_p(u) = 2^{\frac{1}{p}-2} - (1+u^p)^{\frac{1}{p}-2} \cdot u^{p-1} \cdot \left[(p-1) \cdot (1-\sqrt{u}) + \frac{1+u^p}{2}\right] > 0$$

for all $u \in (0,1)$. Because of $\psi_p(1) = 0$ this, however, follows from

$$\psi_p'(u) = -(p-1) \cdot (p - \tfrac{1}{2}) \cdot (1+u^p)^{\frac{1}{p}-3} \cdot u^{p-2} \cdot (1-\sqrt{u}) \cdot (1-u^p) < 0,$$

which is obvious. □

(Received July 18, 1995.)

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. Ser. B *28* (1966), 131–142.

[2] D. E. Boekee: A generalization of the Fisher information measure. Delft University Press, Delft 1977.

[3] I. Csiszár and J. Fischer: Informationsentfernungen im Raum der Wahrscheinlichkeitsverteilungen. Magyar Tud. Akad. Mat. Kutató Int. Kösl. *7* (1962), 159–180.

[4] 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.

[5] D. Feldman and F. Österreicher: Divergenzen von Wahrscheinlichkeitsverteilungen – integralgeometrisch betrachtet. Acta Math. Acad. Sci. Hungar. *37* (1981), 4, 329–337.

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

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

[8] K. Matusita: Decision rules based on the distance for problems of fit, two samples and estimation. Ann. Math. Stat. *26* (1955), 631-640.

[9] F. Österreicher: The construction of least favourable distributions is traceable to a minimal perimeter problem. Studia Sci. Math. Hungar. *17* (1982), 341–351.

[10] F. Österreicher: The risk set of a testing problem – A vivid statistical tool. In: Trans. of the Eleventh Prague Conference, Academia, Prague 1992, Vol. A, pp. 175–188.

[11] E. Reschenhofer and I. M. Bomze: Length tests for goodness of fit. Biometrika *78* (1991), 207–216.

[12] I. Vajda and F. Österreicher: Statistical information and discrimination. IEEE Trans. Inform. Theory *39* (1993), 3, 1036–1039.

[13] I. Vincze: On the concept and measure of information contained in an observation. In: Contributions to Probability (J. Gani and V.F. Rohatgi, eds.), Academic Press, New York 1981, pp. 207–214.

Dr. Ferdinand Österreicher, Institut für Mathematik, Universität Salzburg, Hellbrunnerstraße 34, 5020 Salzburg. Austria.