Kybernetika

Václav Kůs; Domingo Morales; Jitka Hrabáková; Iva Frýdlová

Existence, Consistency and computer simulation for selected variants of minimum distance estimators

Kybernetika, Vol. 54 (2018), No. 2, 336–350

Persistent URL: `http://dml.cz/dmlcz/147198`

KYBERNETIKA — VOLUME 54 (2018), NUMBER 2, PAGES 336–350

EXISTENCE, CONSISTENCY AND COMPUTER SIMULATION FOR SELECTED VARIANTS OF MINIMUM DISTANCE ESTIMATORS

Václav Kůs, Domingo Morales, Jitka Hrabáková and Iva Frýdlová

The paper deals with sufficient conditions for the existence of general approximate minimum distance estimator (AMDE) of a probability density function f_0 on the real line. It shows that the AMDE always exists when the bounded ϕ-divergence, Kolmogorov, Lévy, Cramér, or discrepancy distance is used. Consequently, $n^{-1/2}$ consistency rate in any bounded ϕ-divergence is established for Kolmogorov, Lévy, and discrepancy estimators under the condition that the degree of variations of the corresponding family of densities is finite. A simulation experiment empirically studies the performance of the approximate minimum Kolmogorov estimator (AMKE) and some histogram-based variants of approximate minimum divergence estimators, like power type and Le Cam, under six distributions (Uniform, Normal, Logistic, Laplace, Cauchy, Weibull). A comparison with the standard estimators (moment/maximum likelihood/median) is provided for sample sizes $n = 10, 20, 50, 120, 250$. The simulation analyzes the behaviour of estimators through different families of distributions. It is shown that the performance of AMKE differs from the other estimators with respect to family type and that the AMKE estimators cope more easily with the Cauchy distribution than standard or divergence based estimators, especially for small sample sizes.

Keywords: Kolmogorov distance, ϕ-divergence, minimum distance estimator, consistency rate, computer simulation

Classification: 62B05, 62H30

1. INTRODUCTION AND BASIC CONCEPTS

Minimum distance estimators are increasingly being used when the classical maximum likelihood theory breaks down (unbounded likelihood problems such as mixtures of continuous distributions, heavy tailed distributions with unknown location and scale parameters, or distributions with a parameter dependent support) and because they have good robustness properties. Let us start with basic notations and definitions.

Let $\mathfrak{F}(\mathbb{R})$ be the set of all cumulative distribution functions (cdf) on $(\mathbb{R}, \mathcal{B})$ with the σ-algebra $\mathcal{B}$ of Borel subsets, and $\mathfrak{F} \subset \mathfrak{F}(\mathbb{R})$ be a family of distributions dominated by a σ-finite measure λ on $(\mathbb{R}, \mathcal{B})$. We denote by $\mathfrak{D}$ the set of corresponding probability density functions (pdf) in the Banach space $\mathrm{L}_1(\lambda)$ containing the densities of all distributions

DOI: 10.14736/kyb-2018-2-0336

from $\mathfrak{F}$. Further, $\mathbf{X} = (X_1, \ldots, X_n)$ denotes a random vector whose components are independent and identically distributed (i.i.d.) from $f_0 \in \mathfrak{D}$.

We say that a sequence of mappings $\widehat{f}_n : \mathbb{R}^{n+1} \mapsto [0, \infty)$ is an *estimator* of $f_0 \in \mathfrak{D}$ if, for every $n \in \mathbb{N}$, $\widehat{f}_n$ is measurable and $\widehat{f}_n(\cdot\,, \mathbf{x}) \in \mathfrak{D}$ for every realizations (observations) $\mathbf{x} \in \mathbb{R}^n$ of random variable $\mathbf{X}$. In the rest of paper, $\widehat{F}_n$ denotes the distribution function corresponding to a density estimate $\widehat{f}_n$ and F_n denotes the *empirical distribution functions* on $\mathbb{R}$, i. e.

$$F_n(x) = F_{n,\mathbf{X}}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}_{(-\infty,x]}(X_i), \qquad n \in \mathbb{N},$$

where $\mathbb{I}_A$ denotes the indicator function of $A \subset \mathbb{R}$.

Let D denote a reflexive distance (disparity, divergence, score,...) on $\mathfrak{F} \times \widetilde{\mathfrak{F}}$ for some $\widetilde{\mathfrak{F}} \subset \mathfrak{F}(\mathbb{R})$. By the reflexivity of D we mean that $D(F, F) = 0$ for $F \in \mathfrak{F} \bigcap \widetilde{\mathfrak{F}}$. In case of $\mathfrak{F} \subset \widetilde{\mathfrak{F}}$, each D defines a pseudoreflexive (reflexivity a.e. μ) distance ρ_D on $\mathfrak{D}$ by $\rho_D(f, g) = D(F, G)$, where $F, G \in \mathfrak{F}$ are the distribution functions corresponding to $f, g \in \mathfrak{D}$.

Definition 1.1. An estimator $\widehat{f}_n$ of $f_0 \in \mathfrak{D}$ is the *approximate minimum D (distance) estimator* (AMDE) if the corresponding distribution estimator $\widehat{F}_n \in \mathfrak{F}$ satisfies the condition

$$D(\widehat{F}_n, F_n) \leq \inf_{F \in \mathfrak{F}} D(F, F_n) + o(n^{-1/2}) \quad a.s. \tag{1}$$

If the $o(n^{-1/2})$ term is omitted then $\widehat{f}_n$ is the *minimum D (distance) estimator* (MDE) of f_0. In the setup of parametric statistical inference, we take $\mathfrak{F} = \{F_\theta : \theta \in \Theta\}$ for a given parameter space $\Theta \subset \mathbb{R}^k$; the AMDE of p.d.f. and c.d.f. take the form $\widehat{f}_n = f_{\widehat{\theta}_n}$ and $\widehat{F}_n = F_{\widehat{\theta}_n}$, respectively; and $\widehat{\theta}_n$ is called AMDE or MDE of true θ.

Further, we say that an estimator $\widehat{f}_n$ of $f_0 \in \mathfrak{D}$ is *consistent* in an arbitrary ρ_D-distance if $\rho_D(\widehat{f}_n, f_0) \to 0$ a.s., or consistent in the expected ρ_D-distance if $\mathbf{E}\rho_D(\widehat{f}_n, f_0) \to 0$. We consider $\widehat{f}_n$ to be *consistent of the order of* $r_n \searrow 0$ in the ρ_D-distance if $\rho_D(\widehat{f}_n, f_0) = O_p(r_n)$, or in the expected ρ_D-distance if $\mathbf{E}\rho_D(\widehat{f}_n, f_0) = O(r_n)$.

If D_1, D_2 are two reflexive distances on $\mathfrak{F} \times \widetilde{\mathfrak{F}}$ then D_1 *dominates* D_2 ($D_1 \succ D_2$) on $\mathfrak{F}$ with respect to $\widetilde{\mathfrak{F}}$ if for every $G \in \widetilde{\mathfrak{F}}$ and every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$D_1(F, G) < \delta \quad \text{implies} \quad D_2(F, G) < \varepsilon \quad \text{for all} \quad F \in \mathfrak{F}.$$

Now, we introduce one of the most general and widely used class of distances called ϕ–divergences (Liese and Vajda [20]).

Definition 1.2. Let $F, G \in \mathfrak{F}(\mathbb{R})$, μ be a σ–finite measure on $(\mathbb{R}, \mathcal{B})$ such that $\{F, G\} \ll \mu$ and f, g denote the corresponding Radon-Nikodym derivatives (densities) w.r.t. μ. Then *ϕ–divergence* of F and G is defined by

$$D_\phi(F, G) = \int_{\mathcal{X}} g\, \phi\left(\frac{f}{g}\right) \mathrm{d}\mu, \tag{2}$$

where $\phi : (0,\infty) \to \mathbb{R}$, $\phi(1) = 0$, $\phi(t)$ is convex on $(0,\infty)$ and strictly convex at $t = 1$. We put $g\,\phi(f/g) = g\,\phi(0)$ if $f = 0$ and $g\,\phi(f/g) = f\,\phi(\infty)/\infty$ if $g = 0$, where $\phi(0) := \lim_{t\to 0_+} \phi(t)$ and $\phi(\infty)/\infty := \lim_{t\to\infty} \phi(t)/t$ with the convention '$0 \cdot \infty = 0$'.

The value of $D_\phi(F,G)$ does not depend on the choice of dominating measure μ. It is known, see Csiszár [6, 7], that D_ϕ are all reflexive, $\phi(0) + \phi(\infty)/\infty > 0$, and the range of $D_\phi(F,G)$ is

$$0 \le D_\phi(F,G) \le \phi(0) + \phi(\infty)/\infty, \qquad F,G \in \mathfrak{F}(\mathbb{R}),$$

where the upper bound is achieved if F,G are two singular distributions.

Further, $D_\phi(F,G)$ are all invariant with respect to the linear transformation $\widetilde{\phi}(t) = \phi(t) - \phi'_+(1)(t-1)$, $t \in (0,\infty)$, where $\phi'_+(1)$ denotes derivative of ϕ at $t = 1$ from the right. Thus every ϕ-function has its nonnegative version $\widetilde{\phi}$ producing the same ϕ–divergence while $\widetilde{\phi}'_+(1) = 0$. For details on ϕ-divergences see Vajda [26], Liese and Vajda [21] or Pardo [24]. For the corresponding approximate minimum ϕ–divergence estimator we use the notation AMD$_\phi$E.

If D_ϕ forms a metric distance on $\mathfrak{F}(\mathbb{R}) \times \mathfrak{F}(\mathbb{R})$ then we yield classical approximate minimum distance estimate $\widehat{f}_n$ of f_0. By Vajda [26], D_ϕ is symmetric if and only if $\widetilde{\phi}(t) = t\widetilde{\phi}(1/t)$, $t \in (0,\infty)$. Every D_ϕ can be symmetrized by introducing $\phi_*(t) = \phi(t) + t\phi(1/t)$, $t \in (0,\infty)$. Corresponding ϕ_*-divergence $D_{\phi_*}(F,G)$ is symmetric with $\phi_*(0) + \phi_*(\infty)/\infty = 2\phi(0) + 2\phi(\infty)/\infty$. By Kafka et al. [15], for a given $\alpha > 0$, the symmetric divergence D_ϕ^α is a metric (it satisfies the triangle inequality) if the function $(1-t^\alpha)^{1/\alpha}/\phi(t)$, $t \in (0,\infty)$, is nonincreasing in the domain $t \in (0,1)$. A necessary condition for symmetric divergence D_ϕ^α to be a metric is the boundedness of D_ϕ in the sense of $\phi(0) + \phi(\infty)/\infty < \infty$.

The well-known examples of such metrics are Hellinger squared metric (minimum Hellinger distance estimators, see Beran [3], Győrfi et al. [11]), total variation (minimum L_1-distance estimates, see Győrfi et al. [10]), Le Cam squared metric distance (cf. Le Cam [19]) or Matusita metric distance (Matusita [22]). As can be seen in Ősterreicher [23], the class of ϕ-divergences contains the infinite parametric family of metric divergences.

In Section 2 we give sufficient conditions for the existence of the approximate minimum distance estimators (AMDE) $\widehat{f}_n$ of a p.d.f. f_0 on the real line. In Section 3 we show that the AMDE always exists when the bounded ϕ-divergence or the Kolmogorov distance are used, and its order of consistency in any bounded divergence is given. In order to illustrate the application of AMDE and to get some insight on their small sample size properties, a simulation experiment dealing with AMDE parameter estimation has been carried out in Section 4. Throughout Sections 3-4 we also consider the Kolmogorov distance on $\mathfrak{F}(\mathbb{R}) \times \mathfrak{F}(\mathbb{R})$ given by $K(F,G) = \sup_{x\in\mathbb{R}} |F(x) - G(x)|$, which is not contained in the class of D_ϕ divergences. We use the notation AMKE for this approximate minimum Kolmogorov estimator.

It is important to point out here, that (1) cannot be used generally for arbitrary distance D, or D_ϕ divergence defined above. If it happens that $D(F,F_n) = \infty$ for all $F \in \mathfrak{F}$, than (1) fails down in defining a reasonable version of the estimator $\widehat{F}_n \in \mathfrak{F}$.

Also, using an unbounded divergence D_ϕ with $\phi(0) + \phi(\infty)/\infty = \infty$, we face similar difficulty in the case of lack of absolute continuity between empirical measure P_n and P_F, the probability measures corresponding to F_n and F, respectively. For those reasons we mainly deal with bounded ϕ-divergences and distances in this paper, as in Theorems 3.2 and 3.4.

2. EXISTENCE OF AMDE IN GENERAL CASE

In this section we extend the existence results of Chapter 6 of Pfanzagl [25], established for parametric M-estimators, to the dominated nonparametric case of AMDE. First, we consider the total variational distance on $\mathfrak{F}(\mathbb{R}) \times \mathfrak{F}(\mathbb{R})$

$$V(F,G) = 2 \sup_{B \in \mathcal{B}} \left| \int_B \mathrm{d}F - \int_B \mathrm{d}G \right|, \qquad F, G \in \mathfrak{F}(\mathbb{R}).$$

Since the supremum can be restricted to the ring generated by the semiring of intervals $[a,b)$ with rational values of a and b, the metric distance V is measurable in $\mathbf{X}$ for all distribution functions F_n, G_n which are measurable in the arguments $x \in \mathbb{R}$ and $\mathbf{X} \in \mathbb{R}^n$. If $f, g \in \mathfrak{D}$ are the densities corresponding to distribution functions F, $G \in \mathfrak{F}$ then $V(F,G) = \rho_V(f,g) = \int_{\mathbb{R}} |f - g| \,\mathrm{d}\lambda$. Since $\mathfrak{F}$ is dominated by a σ–finite measure λ on $(\mathbb{R}, \mathcal{B})$ then, by Berger [4], the metric space $(\mathfrak{F}, V)$ is separable which implies the separability of the metric space $(\mathfrak{D}, \rho_V)$. This means that there exists a countable subset $\mathfrak{D}_0 = \{f_{0k} : k \in \mathbb{N}\} \subset \mathfrak{D}$ which is dense in $(\mathfrak{D}, \rho_V)$. The corresponding set of distribution functions $\mathfrak{F}_0 = \{F_{0k} : k \in \mathbb{N}\} \subset \mathfrak{F}$ is countable and dense in $(\mathfrak{F}, V)$.

Lemma 2.1. Let $\mathfrak{F}$, $\widetilde{\mathfrak{F}} \subset \mathfrak{F}(\mathbb{R})$, $\mathfrak{F} \ll \lambda$ and D be a reflexive distance on $\mathfrak{F} \times \widetilde{\mathfrak{F}}$. If $D \prec V$ on $\mathfrak{F}$ w.r.t. $\widetilde{\mathfrak{F}}$ then for every $\varepsilon > 0$ and every $G \in \widetilde{\mathfrak{F}}$ there exists $F_1 \in \mathfrak{F}_0$ such that $D(F_1, G) < \inf_{F \in \mathfrak{F}} D(F,G) + \varepsilon$.

P r o o f. For a fixed arbitrary $\varepsilon > 0$ and $G \in \widetilde{\mathfrak{F}}$ we define the set

$$M = \left\{ F \in \mathfrak{F} : D(F,G) < \inf_{F \in \mathfrak{F}} D(F,G) + \varepsilon \right\} \subset \mathfrak{F}.$$

By definition of infima we know that $M \neq \emptyset$. Denote $r = \inf_{F \in \mathfrak{F}} D(F,G) + \varepsilon$. Since $D \prec V$ on $\mathfrak{F}$ w.r.t. $\widetilde{\mathfrak{F}}$ means simply that $D_G(F) \triangleq D(F,G)$ is a continuous function in F on the separable metric space $(\mathfrak{F}, V)$, then the set $M = D_G^{-1}\Big((-\infty, r)\Big)$ is an open set in $(\mathfrak{F}, V)$. Hence there exists $F_1 \in \mathfrak{F}_0$ such that $F_1 \in M$, i.e. $\inf_{F \in \mathfrak{F}} D(F,G) \leq D(F_1, G) < r$. □

Lemma 2.2. Let D be a reflexive distance on $\mathfrak{F} \times \mathfrak{E}$, where $\mathfrak{E}$ denotes the set of all possible realizations of the empirical distribution functions $F_{n,\mathbf{X}}$ for a given $n \in \mathbb{N}$. Let us further define $\widetilde{D} : \mathfrak{F} \times \mathbb{R}^n \mapsto \mathbb{R}$ by $\widetilde{D}(F, \mathbf{X}) = D(F, F_{n,\mathbf{X}})$, $F \in \mathfrak{F}$. If it holds that

A1) $\widetilde{D}(F,\mathbf{X})$ is measurable with respect to $\mathbf{X}\in\mathbb{R}^n$ for all $F\in\mathfrak{F}$,

A2) $D\prec V$ on $\mathfrak{F}$ w.r.t. $\mathfrak{E}$,

then $\inf_{F\in\mathfrak{F}}\widetilde{D}(F,\mathbf{X})$ is a measurable function in $(\mathbb{R}^n,\mathcal{B}_n)$ with the Borel σ-algebra $\mathcal{B}_n$.

Proof. By Lemma 2.1, for all $\mathbf{X}\in\mathbb{R}^n$ it holds $\inf_{F\in\mathfrak{F}}\widetilde{D}(F,\mathbf{X})=\inf_{F\in\mathfrak{F}_0}\widetilde{D}(F,\mathbf{X})$ which means that $\inf_{F\in\mathfrak{F}}\widetilde{D}(F,\mathbf{X})$ is measurable in $(\mathbb{R}^n,\mathcal{B}_n)$. □

Theorem 2.3. Let D be a reflexive distance on $\mathfrak{F}\times\mathfrak{E}$ satisfying assumptions A1 and A2 of Lemma 2.2. Then there exists at least one AMDE $\widehat{f}_n$ of $f_0\in\mathfrak{D}$.

Proof. Let $a_n=o(n^{-1/2})$, $a_n>0$. We put for every $\mathbf{X}\in\mathbb{R}^n$

$$k^{(n)}(\mathbf{X})=\inf\left\{k\in\mathbb{N}:\ \widetilde{D}(F_{0k},\mathbf{X})<\inf_{F\in\mathfrak{F}}\widetilde{D}(F,\mathbf{X})+a_n,\ F_{0k}\in\mathfrak{F}_0\right\}.$$

Lemma 2.1 ensures that $k^{(n)}(\mathbf{X})$ is well defined by this formula. For every fixed $k\in\mathbb{N}$

$$\left\{\mathbf{X}\in\mathbb{R}^n: k^{(n)}(\mathbf{X})=k\right\}=A_n\cap B_{n,1}\cap B_{n,2}\cap\cdots\cap B_{n,k-1},$$

where

$$\begin{aligned} A_n &= \left\{\mathbf{X}\in\mathbb{R}^n:\widetilde{D}(F_{0k},\mathbf{X})<\inf_{F\in\mathfrak{F}}\widetilde{D}(F,\mathbf{X})+a_n,\ F_{0k}\in\mathfrak{F}_0\right\},\\ B_{n,j} &= \left\{\mathbf{X}\in\mathbb{R}^n:\widetilde{D}(F_{0j},\mathbf{X})\geq\inf_{F\in\mathfrak{F}}\widetilde{D}(F,\mathbf{X})+a_n,\ F_{0j}\in\mathfrak{F}_0\right\},\end{aligned}$$

$j=1,2,\dots,k-1$. By Lemma 2.2, $\inf_{F\in\mathfrak{F}}\widetilde{D}(F,\mathbf{X})$ is measurable in $\mathbf{X}\in\mathbb{R}^n$. Hence the sets A_n and $B_{n,j}$ are all measurable so that the function $k^{(n)}(\mathbf{X})$ is also measurable. Consequently, the composite functions $\widehat{f}_n(x)=f_{0k^{(n)}(\mathbf{X})}(x)$ are measurable in both variables $(x,\mathbf{X})\in\mathbb{R}^{n+1}$. Moreover, $\widehat{f}_n$ obviously satisfies the inequality (1) for its corresponding distribution function $\widehat{F}_n$. □

The existence of MDE's requires stronger assumptions. The existence of a sequence $\{\widehat{f}_n\}$ is ensured only if $(\mathfrak{D},\rho_V)$ is locally compact. Under this assumption the nonparametric modification of the "measurable selection theorem" (cf. Pfanzagl [25], Chap.6) can be applied to the distance D to prove that there exist measurable versions of $\widehat{f}_n$.

3. EXISTENCE AND CONSISTENCY IN SPECIAL CASES

Let us consider the special case $D(F,G)=D_\phi(F,G)$, or $D(F,G)=K(F,G)$, on $\mathfrak{F}(\mathbb{R})\times\mathfrak{F}(\mathbb{R})$. The next Lemma 3.1 presents a version of Proposition 8.27 in Liese and Vajda [20], with a considerably simpler proof.

Lemma 3.1. Let D_ϕ be a ϕ-divergence on $\mathfrak{F}(\mathbb{R})$. Then for all $F, G \in \mathfrak{F}(\mathbb{R})$ it holds

$$D_\phi(F,G) \le [\phi(0) + \phi(\infty)/\infty] \frac{V(F,G)}{2}.$$

P r o o f. By using the convexity of ϕ we obtain for arbitrary $\tau > 1$

$$\phi(t) \le \begin{cases} \phi(0)\,(1-t) + \phi(1)\,t, & \text{for} \quad 0 \le t \le 1, \\ \dfrac{\tau - t}{\tau - 1}\,\phi(1) + \dfrac{t-1}{\tau-1}\,\phi(\tau), & \text{for} \quad 1 \le t \le \tau. \end{cases}$$

Since $\phi(1) = 0$, this implies

$$\phi(t) \le \begin{cases} \phi(0)\,(1-t), & \text{for} \quad 0 \le t \le 1, \\ (t-1)\lim_{\tau\to\infty} \dfrac{\phi(\tau)}{\tau - 1} = (t-1)\,\phi(\infty)/\infty, & \text{for} \quad 1 \le t < \infty. \end{cases}$$

For $A = \{f \ge g\}$ it follows that

$$\begin{aligned} D_\phi(F,G) &\le \int_{\mathcal{X}-A} g\,\phi(0)\left(1 - \frac{f}{g}\right)\,\mathrm{d}\mu + \int_A g\left(\frac{f}{g} - 1\right)\,\phi(\infty)/\infty\;\mathrm{d}\mu \\ &= \phi(0)\int_{\mathcal{X}-A}(g-f)\;\mathrm{d}\mu + (\phi(\infty)/\infty)\int_A (f-g)\;\mathrm{d}\mu \\ &= \phi(0)\,\frac{V(F,G)}{2} + (\phi(\infty)/\infty)\frac{V(F,G)}{2}. \end{aligned}$$

where $\{F, G\} \ll \mu$ and f, g are the corresponding densities w.r.t. μ. □

Theorem 3.2. If D_ϕ is a bounded ϕ-divergence, i.e. $\phi(0) + \phi(\infty)/\infty < \infty$, then an AMD$_\phi$E $\widehat{f}_n$ of the true density $f_0 \in \mathfrak{D}$ always exists. Also, the AMKE $\widehat{f}_n$ of the true density $f_0 \in \mathfrak{D}$ always exists.

P r o o f. The existence of AMD$_\phi$E $\widehat{f}_n$ follows directly from Theorem 2.3 since Lemma 3.1 ensures that $D_\phi \prec V$ on $\mathfrak{F}$ w.r.t. $\mathfrak{E}$ and thus both the assumptions A1 and A2 of Lemma 2.2 are accomplished. For the case of AMKE, it is sufficient to verify A1 and A2 for the Kolmogorov distance K. Concerning A1, we know that $F_{n,\mathbf{X}}(x)$ is measurable in $\mathbf{X} \in \mathbb{R}^n$ for all $x \in \mathbb{R}$. Consequently, $\sup_{x\in\mathbb{Q}} |F(x) - F_{n,\mathbf{X}}(x)|$ is measurable in $\mathbf{X} \in \mathbb{R}^n$ for all $F \in \mathfrak{F}(\mathbb{R})$. Then $K(F, F_{n,\mathbf{X}}) = \sup_{x\in\mathbb{R}} |F(x) - F_{n,\mathbf{X}}(x)| = \sup_{x\in\mathbb{Q}} |F(x) - F_{n,\mathbf{X}}(x)|$ is also measurable. Assumption A2 is valid in the case of Kolmogorov distance due to the fact that $K(F,G) \le V(F,G)/2$, valid for all $F, G \in \mathfrak{F}(\mathbb{R})$. □

By the Theorem 3.2 we have proved the existence of all the previously considered AMD$_\phi$E for bounded ϕ–divergences, e.g. the ones mentioned in Section 1. Moreover, the same or similar reasoning as we did in the proof of Theorem 3.2 can be employed also for the minimum Lévy (L), Cramér (C), and discrepancy (d) distance estimators investigated in Hrabáková and Kůs [14], which are defined through the metric distances

$$L(F,G) = \inf\{\varepsilon > 0 : G(x-\varepsilon) - \varepsilon \le F(x) \le G(x+\varepsilon) + \varepsilon, \forall x \in \mathbb{R}\}\,, \tag{3}$$

$$C(F,G) = \int_{\mathbb{R}} (F(x) - G(x))^2 \,\mathrm{d}G(x)\,, \tag{4}$$

$$d(F,G) = \sup_{B \in \mathbf{B}} |P(B) - Q(B)|\,, \tag{5}$$

respectively, where $\mathbf{B}$ is the set of all closed intervals in $\mathbb{R}$ (or the set of all closed balls in $\mathbb{R}^s$ for multidimensional case), and P, Q are the probability measures corresponding to the densities f, g. All these metric distances (3), (4), and (5), fulfill the measurability assumption A1 of Lemma 2.2. Moreover, the following inequalities

$$\begin{aligned} &L(F,G) \le K(F,G) \le V(F,G)/2\,,\\ &C(F,G) \le [K(F,G)]^2 \le [V(F,G)]^2/4\,,\\ &d(F,G) \le 2K(F,G) \le V(F,G)\,, \end{aligned}$$

induce accomplishment of the A2 domination assumption of Lemma 2.2 for AMLE, AMCE, and AMdE estimators. Thus the AMLE, AMCE and AMdE always exist. The same reasoning can be applied to the combined Kolmogorov–Cramér estimators from Hrabáková and Kůs [13]. (For some other specific inequalities derived in spaces of probability measures see Gibbs and Su [9]).

Now, we address briefly the consistency of AMKE, AMLE, AMCE and AMdE. In the next Theorem 3.4 we apply the previously derived consistency results from Kůs [17] which uses the concept called *degree of variations* of a family of densities $\mathfrak{D}$.

Definition 3.3. Let $\mathfrak{F} \ll \lambda$ and $\mathfrak{D}$ denote the set of densities corresponding to $\mathfrak{F}$. Let f and g be two pdf's and ν^+ and ν^- be the measures on $(\mathbb{R}, \mathcal{B})$ (the lower and upper variations) with corresponding Radon-Nikodym derivatives $(f-g)^+$ and $(f-g)^-$, respectively. We say that $A \in \mathcal{B}$ *separates* ν^+ *and* ν^- if either $\nu^+(A) = \nu^+(\mathbb{R})$ and $\nu^-(\mathbb{R} - A) = \nu^-(\mathbb{R})$ or $\nu^+(\mathbb{R} - A) = \nu^+(\mathbb{R})$ and $\nu^-(A) = \nu^-(\mathbb{R})$. We define *degree of variation* $DV(\mathfrak{D})$ of the family $\mathfrak{D}$ as

$$DV(\mathfrak{D}) = \sup \Big\{ DV(f,g) :\ f, g \in \mathfrak{D} \Big\}$$

where $DV(f,g) = 0$ if the support $A = \{x \in \mathbb{R} : (f-g)^+ > 0\}$ of the component ν^+ separates ν^+ and ν^- while $\lambda(A) = 0$. Otherwise we set

$$DV(f,g) = \inf \Big\{ m \in \mathbb{N} :\ A = \bigcup_{j=1}^{m} \mathcal{I}_j,\ A \text{ separates } \nu^+,\ \nu^- \Big\},$$

where $\mathcal{I}_1, \dots, \mathcal{I}_m$ are nonvoid intervals in $\mathbb{R}$. If the minimized set is empty, i. e. if there is no m of the required properties, we put $DV(f,g) = \infty$.

In brief, this parameter-free quantity $DV(\mathfrak{D})$ characterizes complexity of the family $\mathfrak{D}$ itself with respect to the maximum number of signum changes in differences of every two arbitrarily chosen densities from $\mathfrak{D}$. The degree of variations covers all the parametric, semi-parametric, and nonparametric families, thus the class $\mathfrak{D}$ need not possess any specific structure in general, and it does not require any continuity or differentiability conditions or bounded supports imposed usually on the classes of densities in order to guarantee some consistency properties of estimators under consideration.

Theorem 3.4. If the degree of variations of a family $\mathfrak{D}$ is finite, i. e. $DV(\mathfrak{D}) < \infty$, then the approximate minimum Kolmogorov, Lévy and discrepancy estimators $\widehat{f}_n$ of $f_0 \in \mathfrak{D}$ are all consistent of the order of $n^{-1/2}$ in an arbitrary bounded ϕ-divergence D_ϕ and in an arbitrary expected bounded ϕ-divergence D_ϕ.

Proof. The existence of AMKE, AMLE and AMdE $\widehat{f}_n$ was already proved above. Kůs [17] proved that, under the condition $DV(\mathfrak{D}) < \infty$, the AMKE $\widehat{f}_n$ is $\sqrt{n}$-consistent in the L_1-norm and in the expected L_1-norm. Further, Hrabáková and Kůs [14] extended the same $\sqrt{n}$-consistency result to AMLE and AMdE. Thus, by the inequality proved in Lemma 3.1, all the AMKE, AMLE, and AMdE, are also $\sqrt{n}$-consistent in arbitrary bounded ϕ-divergence D_ϕ. The same consistency rate in any expected bounded ϕ-divergence D_ϕ is implied by the inequality $\mathbf{E}D_\phi(\widehat{F}_n, F_0) \leq [\phi(0)+\phi(\infty)/\infty]\mathbf{E}V(\widehat{F}_n, F_0)/2$, where $\widehat{F}_n$ and F_0 are the distribution functions corresponding to the densities $\widehat{f}_n$ and f_0 (the existence of both expectations follows simply from the the fact that both $D_\phi(\widehat{F}_n, F_0)$ and $V(\widehat{F}_n, F_0)$ are bounded random variables with respect to $\mathbf{X} \in \mathbb{R}^n$). □

The generalized quantity called *partial degree of variations* was introduced in Hrabáková and Kůs [14] and it enables to extend the $\sqrt{n}$-consistency of AMKE, AMLE, and AMdE, also to the families not satisfying the finiteness assumption $DV(\mathfrak{D}) < \infty$ required in Theorem 3.4. Notice also that the approximate minimum Cramér estimator (AMCE) was not treated by Theorem 3.4 concerning consistency since the required $\sqrt{n}$-consistency of AMCE in the (expected) L_1-norm was not yet established theoretically in the current literature. However, the computer simulation carried out on AMCE in Hrabáková and Kůs [14] strongly indicates its $n^{-1/2}$ consistency rate in the L_1–norm.

Győrfi, Vajda and van der Meulen [12] deal with the parametric version of AMKE $f_{\widehat{\theta}_n}$ of f_θ for any parametric family $\mathfrak{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$. By Theorem 3.2 we have proved the existence of their minimum Kolmogorov distance parameter estimates $\widehat{\theta}_n$ and by Theorem 3.4 we established the $\sqrt{n}$-consistency of their density estimators $f_{\widehat{\theta}_n}$ in arbitrary (expected) bounded ϕ-divergence D_ϕ. To derive also classical Euclidean $\sqrt{n}$-consistency of their AMKE $\widehat{\theta}_n$ by means of the results of this paper we would need to establish stronger domination relations between the Euclidean distance on the parametric space Θ and any bounded ϕ-divergence D_ϕ on $\mathfrak{D}$.

4. COMPUTER SIMULATION – AMDE'S PERFORMANCE

In this section we analyze the performance of approximate minimum divergence and approximate minimum Kolmogorov estimators of f_0. These AMD$_\phi$E and AMKE estimators are compared with standard estimators (S) which are known to possess good statistical properties for each considered family of distributions (see below). Throughout the simulation we consider the Kolmogorov distance K as the representant selected from the set of metric distances $\{K, L, C, d\}$. This is reasonable choice since in our previous simulations, carried out in Hrabáková and Kůs [14], the behaviour of approximate minimum Kolmogorov, Lévy, Cramér, and discrepancy distance estimators was shown to be quite similar in respect to sample sizes n and family of distributions undertaken. Further, in our simulation of this Section, we treat two families of divergences. First one is the *power I_α-divergence family*

$$D_{\phi_\alpha}(F,G) = I_\alpha(F,G) = \frac{1}{\alpha(\alpha-1)}\left(\int f^\alpha g^{1-\alpha}\,\mathrm{d}\mu\ -1\right)$$

defined by means of the convex functions $\phi_\alpha(t) = \frac{t^\alpha - 1}{\alpha(\alpha-1)}$, $t > 0$, $\alpha \in (0,1)$, with the limits $\phi_\alpha(0) = \frac{1}{\alpha(1-\alpha)}$ and $\phi_\alpha(\infty)/\infty = 0$. Thus we treat the bounded divergences of Theorem 3.2. Simulations are carried out for $\alpha = 1/4$, $1/2$ and $3/4$ while we tabulate only the case for $\alpha = 1/2$ in this paper, which means that we deal in fact with the approximate minimum Hellinger distance estimators obtained through the squared metric $H^2(F,G) = \int(\sqrt{f}-\sqrt{g})^2\mathrm{d}\mu$. Corresponding AMHE's are denoted by ϕ_1 in Table 1 and Figure 1.

The second family was selected from the class of *extended Le Cam divergences*

$$D_{\phi_\beta}(F,G) = LC_\beta(F,G) = \frac{1}{2}\int \frac{(f-g)^2}{\beta f + (1-\beta)g}\,\mathrm{d}\mu$$

given by means of the convex functions $\phi_\beta(t) = \frac{1}{2}(t-1)^2/\left[(1-\beta)+\beta t\right]$, $t > 0$, $\beta \in (0,1)$, with the finite limits $\phi_\beta(0) = \frac{1}{2(1-\beta)}$ and $\phi_\beta(\infty)/\infty = \frac{1}{2\beta}$. Due to the fact that D_{ϕ_β} becomes the Pearson divergence $\chi^2(F,G)$ for the extreme value $\beta = 0$ and the Neymann divergence for $\beta = 1$, this extended Le Cam divergence is sometimes called *blended divergence* or the *blend of Pearson and Neyman divergences*, see Kůs [16, 18] and further references ibid. Simulation programs run for the values $\beta = 1/4$, $1/2$ and $3/4$ but here we tabulate only the case for $\beta = 1/2$. These approximate minimum Le Cam distance estimators (AMLCE) are denoted by ϕ_2 in Table 1 and Figure 1.

We consider the following classes of distributions and standard (S) estimators:

Class $\mathfrak{F}_1$: Uniform distributions $U(a,b)$, $a,b \in \mathbb{R}$, $a < b$, with standard moment estimators $\widehat{a}_s = \overline{X}_n - \sqrt{3}s_n$ and $\widehat{b}_s = \overline{X}_n + \sqrt{3}s_n$, where $\overline{X}_n = (1/n)\sum_1^n X_i$ and $s_n^2 = (1/n)\sum_1^n (X_i - \overline{X}_n)^2$.

Class $\mathfrak{F}_2$: Gauss (normal) distributions $N(\mu,\sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$, with standard maximum likelihood estimators (MLE) $\widehat{\mu}_s = \overline{X}_n$ and $\widehat{\sigma}_s^2 = s_n^2$.

Class $\mathfrak{F}_3$: Logistic distributions $Lo(\alpha,\beta)$ with the densities

$$f(x) = \frac{1}{\beta}\,exp\left\{-\frac{x-\alpha}{\beta}\right\}\left(1+exp\left\{-\frac{x-\alpha}{\beta}\right\}\right)^{-2}, \qquad x\in\mathbb{R},\quad \alpha\in\mathbb{R},\quad \beta>0,$$

with standard moment estimators $\widehat{\alpha}_s = \overline{X}_n$ and $\widehat{\beta}_s = \sqrt{3}s_n/\pi$.

Class $\mathfrak{F}_4$: Laplace distributions $La(\alpha, \beta)$ with the densities

$$f(x) = \frac{1}{2\beta} \exp\left\{-\frac{|x-\alpha|}{\beta}\right\}, \qquad x \in \mathbb{R}, \quad \alpha \in \mathbb{R}, \quad \beta > 0,$$

and standard moment estimators $\widehat{\alpha}_s = \overline{X}_n$ and $\widehat{\beta}_s = s_n/\sqrt{2}$.

Class $\mathfrak{F}_5$: Cauchy distributions $C(u, v)$ with

$$f(x) = \frac{1}{\pi} \frac{v}{v^2 + (x-u)^2}, \qquad x \in \mathbb{R}, \quad u \in \mathbb{R}, \quad v > 0.$$

Standard estimators are $\widehat{u}_s = \mathbf{X}_{0.5}$ and $\widehat{v}_s = (\mathbf{X}_{0.75} - \mathbf{X}_{0.25})/2$, i.e. the sample median and semi-interquartile range.

Class $\mathfrak{F}_6$: Weibull distributions $W(m, c, b)$,

$$f(x) = \frac{b}{c^b}(x-m)^{b-1} \exp\left\{-\left(\frac{x-m}{c}\right)^b\right\}, \qquad x \geq m, \ m \in \mathbb{R}, \ b > 0, \ c > 0,$$

with standard MLE estimators, $\widehat{m}_s$, $\widehat{c}_s$, $\widehat{b}_s$, obtained by solving the likelihood equations with quantile seeds

$$\widehat{b}_0 = \frac{\ln(-\ln 0.0263) - \ln(-\ln 0.8327)}{\ln X_{([0.9737n])} - \ln X_{([0.1673n])}}, \qquad \widehat{m}_0 = \frac{X_{(1)}X_{(n)} - X_{(2)}^2}{X_{(1)} + X_{(n)} - 2X_{(2)}},$$

where $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote here the ordered sample statistics and $\widehat{c}_s$ is found by direct substitution after the iterative procedure on $\widehat{m}_s$ and $\widehat{b}_s$ has been carried out.

Concerning the simulation experiments, we first generate $L = 10^3$ pseudorandom realizations of the sample $(X_1, X_2, \ldots, X_n)$ i.i.d. F_0, of sizes $n = 10, 20, 50, 120, 250$, taken from one of the above described classes $\mathfrak{F}_1, \ldots, \mathfrak{F}_6$ and we calculate standard (S), AMKE (K), AMD$_{\phi_1}$E (ϕ_1) and AMD$_{\phi_2}$E (ϕ_2) estimates of parameters of the true distribution F_0 (density f_0). We calculate the mean values of these estimates and the mean total variational distance of estimated densities $f_{\widehat{\theta}_{n,l}}$ from the true density f_0, i.e.

$$\overline{V}_n(f_0) := \frac{1}{L}\sum_{l=1}^{L} \rho_V(f_{\widehat{\theta}_{n,l}}, f_0) = \frac{1}{L}\sum_{l=1}^{L} \int_{\mathbb{R}} |f_{\widehat{\theta}_{n,l}} - f_0| \, \mathrm{d}\lambda, \tag{6}$$

where $f_{\widehat{\theta}_{n,l}}$ denotes the parametric density estimate for the selected family of distributions and belonging to the lth repetition, $l = 1, 2, \ldots, L$. We test the performance of the estimators with each of the following source (generating) distributions, F_0, individually: $U(0,1)$, $N(0,1)$, $Lo(0,1)$, $La(0,1)$, $C(0,1)$ and $W(0,1,1)$. Weibull distribution is treated during simulation as 3-parametric with all three parameters unknown and to be estimated by the AMKE and AMD$_\phi$E procedures.

From technical point of view, for the minimization of $D(F_\theta, F_n)$ we applied the numerical method of successively condensed grids combined with gradient method. More precisely, the procedure is the following:

- choose the bounded range of estimated parameters θ,
- lay down the grid with node values,
- find 5 grid nodes with the lowest values of $D(F_\theta, F_n)$,
- for each of these 5 nodes continue in finding of the local minimum by the gradient method,
- obtain five local extremal points,
- the one with the lowest value of $D(F_\theta, F_n)$ becomes the estimate $\widehat{\theta}$ of parameters θ for the considered n, l, $\mathfrak{F}_j$, and D.

For the case of $D = K$, the computation of $K(F_\theta, F_n)$ is simple. For $D = D_{\phi_1}$ and $D = D_{\phi_2}$ we applied classical histogram density estimates $\widehat{f}_n^H$ and set $g = \widehat{f}_n^H$ to evaluate (2). Hence, we work with only approximative (near) AMD$_{\phi_1}$E and AMD$_{\phi_2}$E in our simulation. There are some other approaches how to carry out the evaluating procedure for approximative minimum ϕ–divergence estimates, for example one can apply various Kernel density estimates, Barron density estimates (cf. Barron [2]), or restrict ourselves to only decomposable divergences avoiding the numerical difficulties such as in Frýdlová et al. [8], cf. Broniatowski et al. [5]. Also, the kernel-based minimum dual ϕ-divergence estimator (MD$_\phi$DE) is newly proposed in Al Mohamad [1] (for symmetric and asymmetric kernels), where its asymptotic properties are proved and the efficiency versus robustness is treated through a comprehensive simulation study for two component Gaussian mixture, Weibull mixture, and generalized Pareto distribution. But from practical point of view, we choose our approach based on the well known classical histograms $\widehat{f}_n^H$ to obtain the approximations to AMD$_{\phi_1}$E and AMD$_{\phi_2}$E. Finally, the total variational integral in (6) was then calculated by the Simpson method applied to every interval of the domain of $(f_{\widehat{\theta}_{n,l}} - f_0)$ on which the integrand is continuous.

Table 1 presents the obtained values of $\overline{V}_n(f_0)$. Standard estimates (S) predominantly achieve the best behavior for the normal, logistic and Weibull distributions. However, for the uniform, Laplace and Cauchy distributions (the distributions not satisfying regularity conditions for the asymptotic efficiency of maximum likelihood estimators) one can find minimum distance estimates with better mean L_1–performance that the standard estimates. More precisely, for the uniform distribution, the AMD$_{\phi_1}$E's are significantly best estimates throughout all sample sizes $n = 10$–250, while the AMKE's are the worst ones for small sample sizes $n \leq 50$ and AMD$_{\phi_2}$E's are significantly worst for remaining sample sizes. For the normal distribution, for all $n = 10$–250, the standard estimates are always the best and AMKE's always the worst. In the case of logistic distribution, the situation is exactly the same as for the normal case with the only exception that for the larger sample sizes $n \geq 50$ the AMD$_{\phi_2}$E's are very close (slightly under) to the standard estimates. On the contrary, for both the Laplace and the Cauchy distributions, AMKE's are predominantly the best estimates which can be applied, whilst the AMD$_{\phi_1}$E's do not perform well for small sample sizes. Standard estimates are even the worst for Laplace case if $n \geq 50$. However, within the Weibull family of distributions, the standard estimates attain the lowest mean total variational distances for $n \leq 30$, while AMD$_{\phi_1}$E's slightly overcome them for $n \geq 50$. AMKE's are the worst estimates for Weibull family for all $n = 10$–250.

n	T	$U(0,1)$	$N(0,1)$	$Lo(0,1)$	$La(0,1)$	$C(0,1)$	$W(0,1,1)$
10	S	0.301	0.298	0.325	0.392	0.448	0.401
	K	0.382	0.345	0.346	0.367	0.399	0.482
	ϕ_1	0.150	0.338	0.359	0.405	0.496	0.465
	ϕ_2	0.287	0.336	0.343	0.387	0.437	0.464
20	S	0.206	0.202	0.231	0.284	0.288	0.273
	K	0.256	0.241	0.249	0.264	0.282	0.347
	ϕ_1	0.050	0.239	0.246	0.290	0.355	0.307
	ϕ_2	0.195	0.236	0.239	0.283	0.299	0.328
50	S	0.134	0.128	0.153	0.192	0.182	0.167
	K	0.156	0.154	0.159	0.171	0.174	0.210
	ϕ_1	0.001	0.145	0.157	0.180	0.218	0.165
	ϕ_2	0.154	0.143	0.150	0.172	0.186	0.191
120	S	0.085	0.086	0.097	0.123	0.114	0.095
	K	0.095	0.101	0.101	0.110	0.117	0.135
	ϕ_1	0.000	0.088	0.100	0.118	0.131	0.086
	ϕ_2	0.150	0.087	0.096	0.112	0.124	0.116
250	S	0.060	0.059	0.066	0.085	0.079	0.060
	K	0.068	0.068	0.069	0.076	0.078	0.093
	ϕ_1	0.000	0.063	0.067	0.078	0.092	0.058
	ϕ_2	0.148	0.061	0.065	0.074	0.102	0.066

Tab. 1. Mean total variational distances $\overline{V}_n(f_0)$ for standard (S) and selected variants of AMDE (K, D_{ϕ_1}, D_{ϕ_2}) estimates.

Figure 1 enables to compare more distinctly the behaviour of our estimators through different families of distributions (horizontal directions) and also illustrates their consistency for selected family $\mathfrak{F}$ (vertical directions). Notice that the behaviour of AMKE's differs from the other estimates with respect to family type considered for all sample sizes $n = 10$–250. They achieve minimal L_1–error trend along the normal, logistic and Laplace distributions throughout all the six families $\mathfrak{F}_1, \ldots, \mathfrak{F}_6$. Moreover, the Kolmogorov estimates cope more easily with the Cauchy distribution than standard and AMD$_{\phi_1}$E estimates since Cauchy family is the critical one for them especially for small sample sizes $n \leq 50$. On the contrary, the Weibull is the most problematic class of distributions among $\mathfrak{F}_1, \ldots, \mathfrak{F}_6$, when applying AMKE's or AMD$_{\phi_2}$E's for almost all $n = 10$–250. The increasing horizontal trends in most parts of the graphs for standard, ϕ_1, and ϕ_2, indicates that we have to increase the sample sizes in order to attain the same values of L_1–errors for the distributions from the right-hand part of the graphs. For example, applying AMD$_{\phi_1}$E estimator, for obtaining the same L_1–error around 0.15, we need relatively small sample $n = 10$ observations for the uniform distribution whilst we require moderate sample size around $n = 50$ observations for logistic or Weibull distributions, etc.

If we consider the performances of power divergences I_α for $\alpha = 1/4$, $3/4$, or Le Cam's LC_β for $\beta = 1/4$, $3/4$, we come to the overall conclusion that $I_{1/4}$-based AMDE's, as

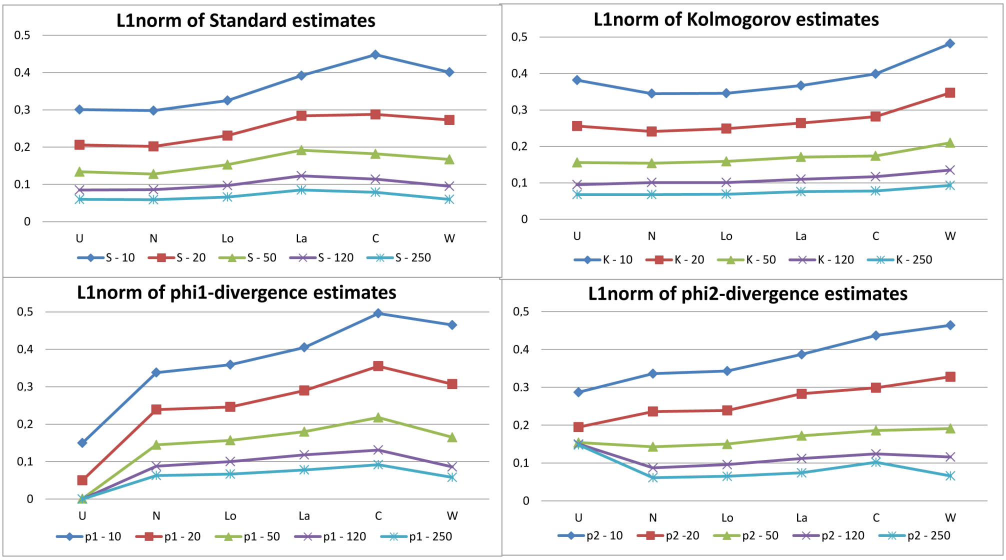

Fig. 1. Mean L_1–errors $\overline{V}_n(f_0)$ for Uniform (U), Normal (N), Logistic (Lo), Laplace (La), Cauchy (C), and Weibull (W) distributions under standard (S), Kolmogorov (K), D_{ϕ_1} (phi1), and D_{ϕ_2} (phi2) AMDE's.

well as $LC_{3/4}$-based AMDE's, preserve the main trends of Figure 1, however they lead to a mild global downgrade of quality (higher L_1-errors), whilst $I_{3/4}$ and $LC_{1/4}$ produce a very slight global improvement of AMDE's in L_1–error, in comparison with D_{ϕ_1} and D_{ϕ_2} estimates reported in Table 1 and Figure 1. The Laplace distribution forms the only exception from this rule for the case of $LC_{1/4}$, where we register significant increase of L_1–error for all sample sizes, even by up to 0.31 for $n = 10$. The downgrade of quality is by around 0.05–0.14 for $I_{1/4}$ divergence estimates and it is by about 0.01–0.05 for $LC_{3/4}$ divergence estimates, depending on distribution family and sample size. The most distinct increase of these L_1-errors was obtained predominantly for small sample sizes, up to $n = 50$, and the worst cases were obtained under the Cauchy distributions. The reason of this L_1 quality drop can be seen in the loss of symmetry of divergences I_α and LC_β, apart from symmetry parameter $\alpha = \beta = 1/2$, and also in the diminished 'weight' $(1 - \beta)$ of the histogram density estimate $g = \widehat{f}_n^H$ in denominator of LC_β. However, further simulations are needed to evaluate convincingly the dependence of AMDE on divergence parameters α and β.

ACKNOWLEDGEMENT

This work was supported by the grants GA16-09848S, SGS15/214/OHK4/3T/14, the Spanish grant MTM2015-64842-P, and partly by LG15047, LM2015068. We also thank for the hospitality of CIO at Miguel Hernández University in Elche (Alicante), where the paper was prepared for publication.

(Received October 19, 2016)

REFERENCES

[1] D. Al Mohamad: Towards a better understanding of the dual representation of phi divergences. Statistical Papers (published on-line 2016.) DOI:10.1007/s00362-016-0812-5

[2] A. R. Barron: The convergence in information of probability density estimators. In: IEEE Int. Symp. Information Theory, Kobe 1988.

[3] R. Beran: Minimum Hellinger distance estimator for parametric models. Ann. Statist. *5* (1977), 455–463. DOI:10.1214/aos/1176343842

[4] A. Berger: Remark on separable spaces of probability measures. An. Math. Statist. *22* (1951), 119–120. DOI:10.1214/aoms/1177729701

[5] M. Broniatowski, A. Toma, and I. Vajda: Decomposable pseudodistances and applications in statistical estimation. J. Statist. Plann. Inference. *142* (2012), 9, 2574–2585. DOI:10.1016/j.jspi.2012.03.019

[6] I. Csiszár: Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizit on Markhoffschen Ketten. Publ. Math. Inst. Hungar. Acad. Sci., Ser. A *8* (1963), 84–108.

[7] I. Csiszár: Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. *2* (1967), 299–318.

[8] I. Frýdlová, I. Vajda, and V. Kůs: Modified power divergence estimators in normal model – simulation and comparative study. Kybernetika *48* (2012), 4, 795–808.

[9] A. L. Gibbs and F. E. Su: On choosing and bounding probability metrics. Int. Statist. Rev. *70* (2002), 419–435. DOI:10.1111/j.1751-5823.2002.tb00178.x

[10] L. Győrfi, I. Vajda, and E. C. van der Meulen: Family of point estimates yielded by L_1-consistent density estimate. In: L_1-Statistical Analysis and Related Methods (Y. Dodge, ed.), Elsevier, Amsterdam 1992, pp. 415–430.

[11] L. Győrfi, I. Vajda, and E. C. van der Meulen: Minimum Hellinger distance point estimates consistent under weak family regularity. Math. Methods Statist. *3* (1994), 25–45.

[12] L. Győrfi, I. Vajda, and E. C. van der Meulen: Minimum Kolmogorov distance estimates of parameters and parametrized distributions. Metrika *43* (1996), 237–255. DOI:10.1007/bf02613911

[13] J. Hrabáková and V. Kůs: The Consistency and Robustness of Modified Cramér-Von Mises and Kolmogorov-Cramér Estimators. Comm. Statist. – Theory and Methods *42* (2013), 20, 3665–3677. DOI:10.1080/03610926.2013.802806

[14] J. Hrabáková and V. Kůs: Notes on consistency of some minimum distance estimators with simulation results. Metrika *80* (2017), 243-257. DOI:10.1007/s00184-016-0601-0

[15] P. Kafka, F. Ősterreicher, and I. Vincze: On powers of f-divergences defining a distance. Studia Sci. Mathem. Hungarica *26* (1991), 415–422.

[16] V. Kůs: Blended ϕ-divergences with examples. Kybernetika *39* (2003), 43-54.

[17] V. Kůs: Nonparametric Density Estimates Consistent of the Order of $n^{-1/2}$ in the L_1-norm. Metrika *60* (2004), 1-14. DOI:10.1007/s001840300286

[18] V. Kůs, D. Morales, and I. Vajda: Extensions of the parametric families of divergences used in statistical inference. Kybernetika *44* (2008), 1, 95–112.

[19] L. Le Cam: Asymptotic Methods in Statistical Decision Theory. Springer, New York 1986. DOI:10.1007/978-1-4612-4946-7

[20] F. Liese and I. Vajda: Convex Statistical Distances. Teubner, Leipzig 1987.

[21] F. Liese and I. Vajda: On divergences and informations in statistics and information theory. IEEE Trans. Inform. Theory *52* (2006), 4394–4412. DOI:10.1109/tit.2006.881731

[22] K. Matusita: Distance and decision rules. Ann. Inst. Statist. Math. *16* (1964), 305-315. DOI:10.1007/bf02868578

[23] F. Ősterreicher: On a class of perimeter-type distances of probability distributions. Kybernetika *32* (1996), 4, 389–393.

[24] L. Pardo: Statistical Inference Based on Divergence Measures. Chapman and Hall, Boston 2006. DOI:10.1201/9781420034813

[25] J. Pfanzagl: Parametric Statistical Theory. W. de Gruyter, Berlin 1994. DOI:10.1515/9783110889765

[26] I. Vajda: Theory of Statistical Inference and Information. Kluwer, Boston 1989.

Václav Kůs, Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2. Czech Republic.
e-mail: vaclav.kus@fjfi.cvut.cz

Domingo Morales, Departamento de Estadística, Matemáticas e Informática, Universidad Miguel Hernández de Elche, Avda. de la Universidad s/n, 03202 Elche (Alicante). España.
e-mail: dmorales@umh.es

Jitka Hrabáková, Department of Applied Mathematics, FIT, Czech Technical University in Prague, Thákurova 9, 160 00 Praha 6. Czech Republic.
e-mail: jitka.hrabakova@fit.cvut.cz

Iva Frýdlová, Statistics and Data Support Department, Czech National Bank, Na Příkopě 28, 115 03 Praha 1. Czech Republic.
e-mail: frydlova.iva@seznam.cz