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

Terms of use:
© Institute of Information Theory and Automation AS CR, 2018

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents
strictly for personal use. Each copy of any part of this document must contain these Terms of use.

This document has been digitized, optimized for electronic delivery and stamped
with digital signature within the project DML-CZ: The Czech Digital Mathematics
Library http://dml.cz
EXISTENCE, CONSISTENCY AND COMPUTER SIMULATION FOR SELECTED VARIANTS OF MINIMUM DISTANCE ESTIMATORS

VÁCLAV KÚS, DOMINGO MORALES, JITKA HRABÁKOVÁ AND IVÁN 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 \( \phi \)-divergence, Kolmogorov, Lévy, Cramér, or discrepancy distance is used. Consequently, \( n^{-1/2} \) consistency rate in any bounded \( \phi \)-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, \( \phi \)-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 \( \mathcal{F}(\mathbb{R}) \) be the set of all cumulative distribution functions (cdf) on \( (\mathbb{R}, \mathcal{B}) \) with the \( \sigma \)-algebra \( \mathcal{B} \) of Borel subsets, and \( \mathcal{F} \subset \mathcal{F}(\mathbb{R}) \) be a family of distributions dominated by a \( \sigma \)-finite measure \( \lambda \) on \( (\mathbb{R}, \mathcal{B}) \). We denote by \( \mathcal{D} \) the set of corresponding probability density functions (pdf) in the Banach space \( L_1(\lambda) \) containing the densities of all distributions.

DOI: 10.14736/kyb-2018-2-0336
from $\mathfrak{g}$. Further, $X = (X_1, \ldots, X_n)$ 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 $\hat{f}_n : \mathbb{R}^{n+1} \mapsto [0, \infty)$ is an estimator of $f_0 \in \mathfrak{D}$ if, for every $n \in \mathbb{N}$, $\hat{f}_n$ is measurable and $\hat{f}_n(\cdot, x) \in \mathfrak{D}$ for every realizations (observations) $x \in \mathbb{R}^n$ of random variable $X$. In the rest of paper, $\hat{F}_n$ denotes the distribution function corresponding to a density estimate $\hat{f}_n$ and $F_n$ denotes the empirical distribution functions on $\mathbb{R}$, i.e.

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

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

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

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

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

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

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

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

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

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

**Definition 1.2.** Let $F, G \in \mathfrak{g}(\mathbb{R})$, $\mu$ be a $\sigma$–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. $\mu$. Then $\phi$–divergence of $F$ and $G$ is defined by

$$D_{\phi}(F, G) = \int_{\mathcal{X}} g \phi \left( \frac{f}{g} \right) \, d\mu, \quad (2)$$
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 $\mu$. It is known, see Csiszár [6, 7], that $D_\phi$ are all reflexive, $\phi(0) + \phi(\infty) / \infty > 0$, and the range of $D_\phi(F, G)$ is

$$0 \leq D_\phi(F, G) \leq \phi(0) + \phi(\infty) / \infty, \quad F, G \in \mathcal{F}(\mathbb{R}),$$

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 $\tilde{\phi}(t) = \phi(t) - \phi'(1)(t-1)$, $t \in (0, \infty)$, where $\phi'(1)$ denotes derivative of $\phi$ at $t = 1$ from the right. Thus every $\phi$-function has its nonnegative version $\tilde{\phi}$ producing the same $\phi$-divergence while $\phi'(1) = 0$. For details on $\phi$-divergences see Vajda [26], Liese and Vajda [21] or Pardo [24]. For the corresponding approximate minimum $\phi$-divergence estimator we use the notation AMD$E$.

If $D_\phi$ forms a metric distance on $\mathcal{F}(\mathbb{R}) \times \mathcal{F}(\mathbb{R})$ then we yield classical approximate minimum distance estimate $\tilde{f}_n$ of $f_0$. By Vajda [26], $D_\phi$ is symmetric if and only if $\tilde{\phi}(t) = t\tilde{\phi}(1/t)$, $t \in (0, \infty)$. Every $D_\phi$ can be symmetrized by introducing $\phi_+(t) = \phi(t) + t\phi(1/t)$, $t \in (0, \infty)$. Corresponding $\phi_+$-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 $D_\alpha$ 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 $D_\alpha$ to be a metric is the boundedness of $D_\phi$ 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 Österreicher [23], the class of $\phi$-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) $\tilde{f}_n$ of a p.d.f. $f_0$ on the real line. In Section 3 we show that the AMDE always exists when the bounded $\phi$-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 $\mathcal{F}(\mathbb{R}) \times \mathcal{F}(\mathbb{R})$ given by $K(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|$, which is not contained in the class of $D_\phi$ 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 $D_\phi$ divergence defined above. If it happens that $D(F, F_n) = \infty$ for all $F \in \mathcal{F}$, than [1] fails down in defining a reasonable version of the estimator $\tilde{F}_n \in \mathcal{F}$.
Also, using an unbounded divergence \( D_\phi \) 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 \( \phi \)-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 \( \mathfrak{F}(\mathbb{R}) \times \mathfrak{F}(\mathbb{R}) \)

\[
V(F,G) = 2\sup_{B \in \mathcal{B}} \left| \int_B dF - \int_B dG \right|, \quad F,G \in \mathfrak{F}(\mathbb{R}).
\]

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 \( X \) for all distribution functions \( F_n, G_n \) which are measurable in the arguments \( x \in \mathbb{R} \) and \( X \in \mathbb{R}^n \). If \( f, g \in \mathcal{D} \) are the densities corresponding to distribution functions \( F, G \in \mathfrak{F} \) then \( V(F,G) = \rho_V(f,g) = \int_\mathbb{R} |f - g| d\lambda \). Since \( \mathfrak{F} \) is dominated by a \( \sigma \)-finite measure \( \lambda \) on \((\mathbb{R}, \mathcal{B})\) then, by Berger [4], the metric space \((\mathfrak{F}, \rho_V)\) 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 \( \mathfrak{F}, \mathfrak{F}_0 \subset \mathfrak{F}(\mathbb{R}), \mathfrak{F} \ll \lambda \) and \( D \) be a reflexive distance on \( \mathfrak{F} \times \mathfrak{F}_0 \). If \( D < V \) on \( \mathfrak{F}_0 \) w.r.t. \( \mathfrak{F} \) then for every \( \varepsilon > 0 \) and every \( G \in \mathfrak{F} \) there exists \( F_1 \in \mathfrak{F}_0 \) such that \( D(F_1, G) < \inf_{F \in \mathfrak{F}} D(F, G) + \varepsilon \).

**Proof.** For a fixed arbitrary \( \varepsilon > 0 \) and \( G \in \mathfrak{F}_0 \) we define the set

\[
M = \left\{ F \in \mathfrak{F} : D(F,G) < \inf_{F \in \mathfrak{F}_0} D(F, G) + \varepsilon \right\} \subset \mathfrak{F}.
\]

By definition of infima we know that \( M \neq \emptyset \). Denote \( r = \inf_{F \in \mathfrak{F}} D(F, G) + \varepsilon \). Since \( D < V \) on \( \mathfrak{F}_0 \) w.r.t. \( \mathfrak{F} \) means simply that \( D_G(F) \doteq D(F,G) \) is a continuous function in \( F \) on the separable metric space \((\mathfrak{F}, V)\), then the set \( M = D_G^{-1}\left((-\infty, r]\right) \) 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 \( F_n(X) \) for a given \( n \in \mathbb{N} \). Let us further define \( \tilde{D} : \mathfrak{F} \times \mathbb{R}^n \to \mathbb{R} \) by \( \tilde{D}(F,X) = D(F, F_n(X), F \in \mathfrak{F} \). If it holds that
A1) $\widetilde{D}(F, X)$ is measurable with respect to $X \in \mathbb{R}^n$ for all $F \in \mathcal{F}$,

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

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

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

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

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

$$k^{(n)}(X) = \inf \left\{ k \in \mathbb{N} : \tilde{D}(F_0k, X) < \inf_{F \in \mathcal{E}} \tilde{D}(F, X) + a_n, F_0k \in \mathcal{F}_0 \right\}.$$  

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

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

where

$$A_n = \left\{ X \in \mathbb{R}^n : \tilde{D}(F_0k, X) < \inf_{F \in \mathcal{E}} \tilde{D}(F, X) + a_n, F_0k \in \mathcal{F}_0 \right\},$$

$$B_{n,j} = \left\{ X \in \mathbb{R}^n : \tilde{D}(F_0j, X) \geq \inf_{F \in \mathcal{E}} \tilde{D}(F, X) + a_n, F_0j \in \mathcal{F}_0 \right\},$$

$j = 1, 2, \ldots, k - 1$. By Lemma 2.2 $\inf_{F \in \mathcal{E}} \tilde{D}(F, X)$ is measurable in $X \in \mathbb{R}^n$. Hence the sets $A_n$ and $B_{n,j}$ are all measurable so that the function $k^{(n)}(X)$ is also measurable. Consequently, the composite functions $\hat{f}_n(x) = f_0^{(n)}(X_j)(x)$ are measurable in both variables $(x, X) \in \mathbb{R}^{n+1}$. Moreover, $\hat{f}_n$ obviously satisfies the inequality (1) for its corresponding distribution function $\hat{F}_n$. $\square$

The existence of MDE’s requires stronger assumptions. The existence of a sequence $\{\hat{f}_n\}$ is ensured only if $(\mathcal{D}, \rho_Y)$ 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 $\hat{f}_n$.

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 $\mathcal{F}(\mathbb{R}) \times \mathcal{F}(\mathbb{R})$. 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 $D_\phi$ be a $\phi$-divergence on $\mathcal{F}(\mathbb{R})$. Then for all $F, G \in \mathcal{F}(\mathbb{R})$ it holds

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

Proof. By using the convexity of $\phi$ we obtain for arbitrary $\tau > 1$

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

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

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

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

$$D_\phi(F, G) \leq \int_{X - A} g \phi(0) \left(1 - \frac{f}{g}\right) \, d\mu + \int_A g \left(\frac{f}{g} - 1\right) \phi(\infty)/\infty \, d\mu = \phi(0) \int_{X - A} (g - f) \, d\mu + \phi(\infty)/\infty \int_A (f - g) \, d\mu = \phi(0) \frac{V(F, G)}{2} + \phi(\infty)/\infty \frac{V(F, G)}{2},$$

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

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

Proof. The existence of AMD$_\phi E \hat{f}_n$ follows directly from Theorem 2.3 since Lemma 3.1 ensures that $D_\phi \prec V$ on $\mathcal{F}$ w.r.t. $\mathcal{C}$ 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 $F_{n, X}(x)$ is measurable in $X \in \mathbb{R}^n$ for all $x \in \mathbb{R}$. Consequently, $\sup_{x \in \mathbb{Q}} |F(x) - F_{n, X}(x)|$ is measurable in $X \in \mathbb{R}^n$ for all $F \in \mathcal{F}(\mathbb{R})$. Then $K(F, F_{n, X}) = \sup_{x \in \mathbb{R}} |F(x) - F_{n, X}(x)| = \sup_{x \in \mathbb{Q}} |F(x) - F_{n, X}(x)|$ is also measurable. Assumption A2 is valid in the case of Kolmogorov distance due to the fact that $K(F, G) \leq V(F, G)/2$, valid for all $F, G \in \mathcal{F}(\mathbb{R})$. \qed

By the Theorem 3.2 we have proved the existence of all the previously considered AMD$_\phi E$ for bounded $\phi$-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

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

\[ C(F,G) = \int_{\mathbb{R}} (F(x) - G(x))^2 \, dG(x), \tag{4} \]

\[ d(F,G) = \sup_{B \in \mathcal{B}} |P(B) - Q(B)|, \tag{5} \]

respectively, where \( \mathcal{B} \) 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

\[ L(F,G) \leq K(F,G) \leq V(F,G)/2, \]

\[ C(F,G) \leq [K(F,G)]^2 \leq [V(F,G)]^2/4, \]

\[ d(F,G) \leq 2K(F,G) \leq V(F,G), \]

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 \( \mathcal{D} \).

**Definition 3.3.** Let \( \mathfrak{F} \ll \lambda \) and \( \mathcal{D} \) denote the set of densities corresponding to \( \mathfrak{F} \). Let \( f \) and \( g \) be two pdf's and \( \nu^+ \) and \( \nu^- \) 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 \( \nu^+ \) and \( \nu^- \) if either \( \nu^+(\mathbb{R}) = \nu^+(\mathbb{R}) \) 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 \( DV(\mathcal{D}) \) of the family \( \mathcal{D} \) as

\[ DV(\mathcal{D}) = \sup \{ DV(f,g) : f, g \in \mathcal{D} \} \]

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

\[ DV(f,g) = \inf \{ m \in \mathbb{N} : A = \bigcup_{j=1}^{m} I_j, \ A \ \text{separates} \ \nu^+, \nu^- \}, \]

where \( I_1, \ldots, I_m \) 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 $DV(\mathcal{D})$ characterizes complexity of the family $\mathcal{D}$ itself with respect to the maximum number of signum changes in differences of every two arbitrarily chosen densities from $\mathcal{D}$. The degree of variations covers all the parametric, semi-parametric, and nonparametric families, thus the class $\mathcal{D}$ 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 $\mathcal{D}$ is finite, i.e. $DV(\mathcal{D}) < \infty$, then the approximate minimum Kolmogorov, Lévy and discrepancy estimators $\hat{f}_n$ of $f_0 \in \mathcal{D}$ are all consistent of the order of $n^{-1/2}$ in an arbitrary bounded $\phi$-divergence $D_\phi$ and in an arbitrary expected bounded $\phi$-divergence $D_\phi$.

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

The generalized quantity called **partial degree of variations** was introduced in Hrabáková and Kús [14] and it enables to extend the $\sqrt{n}$-consistency of AMKE, AMLE, and AMdE, also to the families not satisfying the finiteness assumption $DV(\mathcal{D}) < \infty$ 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 Kús [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 $\hat{f}_{\hat{\theta}}_n$ of $f_\theta$ for any parametric family $\mathcal{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$. By Theorem 3.2 we have proved the existence of their minimum Kolmogorov distance parameter estimates $\hat{\theta}_n$ and by Theorem 3.4 we established the $\sqrt{n}$-consistency of their density estimators $\hat{f}_{\hat{\theta}}_n$ in arbitrary (expected) bounded $\phi$-divergence $D_\phi$. To derive also classical Euclidean $\sqrt{n}$-consistency of their AMKE $\hat{\theta}_n$ by means of the results of this paper we would need to establish stronger domination relations between the Euclidean distance on the parametric space $\Theta$ and any bounded $\phi$-divergence $D_\phi$ on $\mathcal{D}$.

**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$D_\phi$E and AMKE esti-
mators 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 $I_{\alpha}$-divergence family

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

defined by means of the convex functions $\phi_{\alpha}(t) = \frac{t^\alpha - 1}{\alpha(1-\alpha)}$, $t > 0$, $\alpha \in (0,1)$, 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 $H^2(F,G) = \int \sqrt{f} - \sqrt{g} \, d\mu$. Corresponding AMHE’s are denoted by $\phi_1$ in Table 1 and Figure 1.

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

$$D_{\phi_{\beta}}(F,G) = LC_{\beta}(F,G) = \frac{1}{2} \int \frac{(f-g)^2}{\beta f + (1-\beta)g} \, d\mu$$
given by means of the convex functions $\phi_{\beta}(t) = \frac{1}{2}(t-1)^2 / [(1 - \beta) + \beta t]$, $t > 0$, $\beta \in (0,1)$, 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 $D_{\phi_{\beta}}$ 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 $\phi_2$ 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 $\hat{\theta}_s = \overline{X}_n - \sqrt{3}s_n$ and $\hat{s}_s = \overline{X}_n + \sqrt{3}s_n$, 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) $\hat{\theta}_s = \overline{X}_n$ and $\hat{s}_s^2 = s_n^2$.

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

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

with standard moment estimators $\hat{\theta}_s = \overline{X}_n$ and $\hat{s}_s = \sqrt{3}s_n/\pi$. 
Class $\mathcal{F}_4$: Laplace distributions $La(\alpha, \beta)$ with the densities
\[
f(x) = \frac{1}{2\beta} \exp\left\{ -\frac{|x-\alpha|}{\beta} \right\}, \quad x \in \mathbb{R}, \ \alpha \in \mathbb{R}, \ \beta > 0,
\]
and standard moment estimators $\hat{\alpha}_s = \bar{X}_n$ and $\hat{\beta}_s = s_n/\sqrt{2}$.

Class $\mathcal{F}_5$: Cauchy distributions $C(u, v)$ with
\[
f(x) = \frac{1}{\pi v^2 + (x-u)^2}, \quad x \in \mathbb{R}, \ u \in \mathbb{R}, \ v > 0.
\]
Standard estimators are $\hat{u}_s = X_{0.5}$ and $\hat{v}_s = (X_{0.75} - X_{0.25})/2$, i.e. the sample median and semi-interquartile range.

Class $\mathcal{F}_6$: Weibull distributions $W(m, c, b)$,
\[
f(x) = \frac{b}{c^b} (x-m)^{b-1} \exp\left\{ -\left(\frac{x-m}{c}\right)^b \right\}, \quad x \geq m, \ m \in \mathbb{R}, \ b > 0, \ c > 0,
\]
with standard MLE estimators, $\hat{m}_s$, $\hat{c}_s$, $\hat{b}_s$, obtained by solving the likelihood equations with quantile seeds
\[
\hat{b}_0 = \ln(-\ln 0.0263) - \ln(-\ln 0.8327) = \frac{\ln X_{(0.9737n)} - \ln X_{(0.1673n)}}{\ln X_{(0.9737n)} - \ln X_{(0.1673n)}}, \quad \hat{c}_0 = \frac{X_{(1)}X_{(n)} - X_{(2)}^2}{X_{(1)} + X_{(n)} - 2X_{(2)}},
\]
where $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote here the ordered sample statistics and $\hat{c}_s$ is found by direct substitution after the iterative procedure on $\hat{m}_s$ and $\hat{b}_s$ 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 $\mathcal{F}_1, \ldots, \mathcal{F}_6$ and we calculate standard $(S)$, AMKE $(K)$, AMD$\phi_1 E$ $(\phi_1)$ and AMD$\phi_2 E$ $(\phi_2)$ 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 $f_{\hat{\theta}_{n,l}}$ from the true density $f_0$, i.e.

\[
\overline{V}_n(f_0) := \frac{1}{L} \sum_{l=1}^L \rho_V(f_{\hat{\theta}_{n,l}}, f_0) = \int_{\mathbb{R}} |f_{\hat{\theta}_{n,l}} - f_0| \, d\lambda,
\]  

where $f_{\hat{\theta}_{n,l}}$ denotes the parametric density estimate for the selected family of distributions and belonging to the $l$th 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 $D(F_0, F_n)$ 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 $\theta$,
• 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 $\hat{\theta}$ of parameters $\theta$ for the considered $n$, $l$, $j$, 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 $\hat{f}_n^H$ and set $g = \hat{f}_n^H$ 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 $\phi$–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 $\phi$-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 $\hat{f}_n^H$ 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 $(\hat{f}_n^\theta - f_0)$ on which the integrand is continuous.

Table 1 presents the obtained values of $V_n (f_0)$. 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 $n = 10–250$. 
Existence, consistency and compute simulations for MDE’s

<table>
<thead>
<tr>
<th>$n$</th>
<th>$T$</th>
<th>$U(0, 1)$</th>
<th>$N(0, 1)$</th>
<th>$Lo(0, 1)$</th>
<th>$La(0, 1)$</th>
<th>$C(0, 1)$</th>
<th>$W(0, 1, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>$S$</td>
<td>0.301</td>
<td>0.298</td>
<td>0.325</td>
<td>0.392</td>
<td>0.448</td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td>$K$</td>
<td>0.382</td>
<td>0.345</td>
<td>0.346</td>
<td>0.367</td>
<td>0.399</td>
<td>0.482</td>
</tr>
<tr>
<td></td>
<td>$\phi_1$</td>
<td>0.150</td>
<td>0.338</td>
<td>0.359</td>
<td>0.405</td>
<td>0.496</td>
<td>0.465</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>0.287</td>
<td>0.336</td>
<td>0.343</td>
<td>0.387</td>
<td>0.437</td>
<td>0.464</td>
</tr>
<tr>
<td>20</td>
<td>$S$</td>
<td>0.206</td>
<td>0.202</td>
<td>0.231</td>
<td>0.284</td>
<td>0.288</td>
<td>0.273</td>
</tr>
<tr>
<td></td>
<td>$K$</td>
<td>0.256</td>
<td>0.241</td>
<td>0.249</td>
<td>0.264</td>
<td>0.282</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td>$\phi_1$</td>
<td>0.050</td>
<td>0.239</td>
<td>0.246</td>
<td>0.290</td>
<td>0.355</td>
<td>0.307</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>0.195</td>
<td>0.236</td>
<td>0.239</td>
<td>0.283</td>
<td>0.299</td>
<td>0.328</td>
</tr>
<tr>
<td>50</td>
<td>$S$</td>
<td>0.134</td>
<td>0.128</td>
<td>0.153</td>
<td>0.192</td>
<td>0.182</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>$K$</td>
<td>0.156</td>
<td>0.154</td>
<td>0.159</td>
<td>0.171</td>
<td>0.174</td>
<td>0.210</td>
</tr>
<tr>
<td></td>
<td>$\phi_1$</td>
<td>0.001</td>
<td>0.145</td>
<td>0.157</td>
<td>0.180</td>
<td>0.218</td>
<td>0.165</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>0.154</td>
<td>0.143</td>
<td>0.150</td>
<td>0.172</td>
<td>0.186</td>
<td>0.191</td>
</tr>
<tr>
<td>120</td>
<td>$S$</td>
<td>0.085</td>
<td>0.086</td>
<td>0.097</td>
<td>0.123</td>
<td>0.114</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>$K$</td>
<td>0.095</td>
<td>0.101</td>
<td>0.101</td>
<td>0.110</td>
<td>0.117</td>
<td>0.135</td>
</tr>
<tr>
<td></td>
<td>$\phi_1$</td>
<td>0.000</td>
<td>0.088</td>
<td>0.100</td>
<td>0.118</td>
<td>0.131</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>0.150</td>
<td>0.087</td>
<td>0.096</td>
<td>0.112</td>
<td>0.124</td>
<td>0.116</td>
</tr>
<tr>
<td>250</td>
<td>$S$</td>
<td>0.060</td>
<td>0.059</td>
<td>0.066</td>
<td>0.085</td>
<td>0.079</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>$K$</td>
<td>0.068</td>
<td>0.068</td>
<td>0.069</td>
<td>0.076</td>
<td>0.078</td>
<td>0.093</td>
</tr>
<tr>
<td></td>
<td>$\phi_1$</td>
<td>0.000</td>
<td>0.063</td>
<td>0.067</td>
<td>0.078</td>
<td>0.092</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>$\phi_2$</td>
<td>0.148</td>
<td>0.061</td>
<td>0.065</td>
<td>0.074</td>
<td>0.102</td>
<td>0.066</td>
</tr>
</tbody>
</table>

**Tab. 1.** Mean total variational distances $V_n(f_0)$ for standard $(S)$ and selected variants of AMDE $(K, D_{\phi_1}, D_{\phi_2})$ 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 $F$ (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 $F_1, \ldots, F_6$. Moreover, the Kolmogorov estimates cope more easily with the Cauchy distribution than standard and $AMD_{\phi_1E}$ 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 $F_1, \ldots, F_6$, when applying AMKE’s or $AMD_{\phi_2E}$’s for almost all $n = 10–250$. The increasing horizontal trends in most parts of the graphs for standard, $\phi_1$, and $\phi_2$, 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 $AMD_{\phi_1E}$ 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 $I_{\alpha}$ for $\alpha = 1/4, 3/4$, or Le Cam’s $LC_{\beta}$ 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 $\mathbf{V}_n(f_0)$ for Uniform (U), Normal (N), Logistic (Lo), Laplace (La), Cauchy (C), and Weibull (W) distributions under standard (S), Kolmogorov (K), $D_{\phi_1}$ (phi1), and $D_{\phi_2}$ (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 $D_{\phi_1}$ and $D_{\phi_2}$ 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 $I_\alpha$ and $LC_\beta$, apart from symmetry parameter $\alpha = \beta = 1/2$, and also in the diminished ‘weight’ $(1 - \beta)$ of the histogram density estimate $g = \hat{f}_n^H$ in denominator of $LC_\beta$. However, further simulations are needed to evaluate convincingly the dependence of AMDE on divergence parameters $\alpha$ and $\beta$.

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.
REFERENCES


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