GENERAL BOOTSTRAP FOR DUAL ϕ-DIVERGENCES ESTIMATES

SALIM BOUZEBDA AND MOHAMED CHERFI

ABSTRACT. A general notion of bootstrapped ϕ-divergences estimates constructed by exchangeably weighting sample is introduced. Asymptotic properties of these generalized bootstrapped ϕ-divergences estimates are obtained by using the empirical process theory. Some of practical problems are discussed. Simulation results are used to illustrate the finite sample performance of the proposed estimators.

AMS Subject Classification : 62F40 ; 62F35 ; 62F12; 62G20 ; 62G09 ; 62G30.
Key words and phrases : Bootstrap consistency ; weighted bootstrap ; Kaplan-Meier ; parametric model ; M-estimators ; ϕ-divergence.

1. Introduction

The ϕ-divergence modeling has proved to be a flexible and provided a powerful statistical modeling framework in a variety of applied and theoretical contexts [see Broniatowski and Keziou (2009), Pardo (2006) and Liese and Vajda (2006, 1987)]. Unfortunately, in general, the limiting distribution of the estimators or their functionals based on ϕ-divergences depend crucially on the unknown distribution which is a serious problem in practice. To circumvent this matter, we shall propose, in this work, a general bootstrap of ϕ-divergence based estimators and study some of its properties by mean of a sophisticated empirical process techniques.

A major application for an estimator is in the calculation of confidence intervals. By far the most favored confidence interval is the standard confidence interval based on a normal or a Student t-distribution. Such standard intervals are useful tools, but they are based on an approximation that can be quite inaccurate in practice. Bootstrap procedures are an attractive alternative. One way to look at them is as procedures for handling data when one is not willing to make assumptions about the parameters of the populations from which one sampled. The most that one is willing to assume is that the data are a reasonable representation of the population from

which they came. One then resamples from the data and draws inferences about the corresponding population and its parameters. The resulting confidence intervals have received the most theoretical study of any topic in the bootstrap analysis.
Our main findings, which are analogous to that of Cheng and Huang (2010), are summarized as follows. The ϕ-divergence estimator $\widehat{\boldsymbol{\alpha}}_{\phi}(\boldsymbol{\theta})$ and the bootstrap ϕ-divergence estimator $\widehat{\boldsymbol{\alpha}}^{*}_{\phi}(\boldsymbol{\theta})$ are obtained by optimizing the objective function $h(\boldsymbol{\theta},\boldsymbol{\alpha})$ based on the independent and identically distributed [i.i.d.] observations $X_1,\ldots,X_n$ and the bootstrap sample $X_1^*,\ldots,X_n^*$, respectively,

$$\widehat{\boldsymbol{\alpha}}_{\phi}(\boldsymbol{\theta}) := \arg\sup_{\boldsymbol{\alpha}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{\theta},\boldsymbol{\alpha},X_i), \tag{1.1}$$

$$\widehat{\boldsymbol{\alpha}}^{*}_{\phi}(\boldsymbol{\theta}) := \arg\sup_{\boldsymbol{\alpha}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{i=1}^{n} h(\boldsymbol{\theta},\boldsymbol{\alpha},X_i^*), \tag{1.2}$$

where $X_1^*,\ldots,X_n^*$ are independent draws with replacement from the original sample. We shall mention that $\widehat{\boldsymbol{\alpha}}^{*}_{\phi}(\boldsymbol{\theta})$ can alternatively be expressed as

$$\widehat{\boldsymbol{\alpha}}^{*}_{\phi}(\boldsymbol{\theta}) = \arg\sup_{\boldsymbol{\alpha}\in\boldsymbol{\Theta}} \frac{1}{n}\sum_{i=1}^{n} W_{ni} h(\boldsymbol{\theta},\boldsymbol{\alpha},X_i) \tag{1.3}$$

where the bootstrap weights

$$(W_{n1},\ldots,W_{nn}) \sim \quad \text{Multinomial}(n; n^{-1},\ldots,n^{-1}).$$

In this paper, we shall consider the more general exchangeable bootstrap weighting scheme that includes Efron's bootstrap [Efron (1979) and Efron and Tibshirani (1993)]. The general resampling scheme was first proposed in Rubin (1981) and extensively studied by Bickel and Freedman (1981), who suggested the name "weighted bootstrap", e.g., Bayesian Bootstrap when $(W_{n1},\ldots,W_{nn}) = (D_{n1},\ldots,D_{nn})$ is equal in distribution to the vector of n spacings of $n-1$ ordered uniform $(0,1)$ random variables, that is

$$(D_{n1},\ldots,D_{nn}) \sim \quad \text{Dirichlet}(n;1,\ldots,1).$$

The interested reader may refer to Lo (1993). The case

$$(D_{n1},\ldots,D_{nn}) \sim \quad \text{Dirichlet}(n;4,\ldots,4)$$

was considered in Weng (1989, Remark 2.3) and Zheng and Tu (1988, Remrak 5.) The Bickel and Freedman result concerning the empirical process has been subsequently generalized for empirical processes based on observations in $\mathbb{R}^d$, $d > 1$ as well as in very general sample spaces and for various set and function-indexed random objects [see, for example Beran (1984), Beran and Millar (1986), Beran *et al.* (1987), Gaenssler (1992), Lohse (1987)]. In this setting, Csörgő and Mason (1989) developed similar results for a variety of other statistical functions. This line of research was continued in the work of Giné and Zinn (1989, 1990). There is a huge literature on the application of the bootstrap methodology to nonparametric kernel density and regression estimation, among other statistical procedures, and it is not the purpose of this paper to survey this extensive literature. This being said, it is worthwhile mentioning that the bootstrap as per Efron's original formulation (see Efron (1979)) presents some drawbacks. Namely, some observations may be used more than once while others are not sampled at all. To overcome this difficulty, a more general formulation of the bootstrap has been devised: the *weighted* (or *smooth*) bootstrap, which has also been shown to be computationally more efficient in several applications. We may refer to Mason and Newton (1992), Præstgaard and Wellner (1993) and del Barrio and Matrán (2000). Holmes and Reinert (2004) provided new proofs for many known results about the convergence in law of the bootstrap distribution to the true distribution of smooth statistics employing the techniques based on Stein's method for empirical processes. Note that other variations of Efron's bootstrap are studied in Chatterjee and Bose (2005) using the term "generalized bootstrap". The practical usefulness of the more general scheme is well-documented in the literature. For a survey of further results on weighted bootstrap the reader is referred to Barbe and Bertail (1995).

The remainder on this paper is organized as follows. In the forthcoming section we recall the estimation procedure based on ϕ-divergences and its bootstrap is introduced in details. The asymptotic properties of bootstrapped estimators are given. In section 3, we illustrate how to apply our results in the context of right censoring. Section 4 provides simulation results in order to illustrate the performance of the proposed estimators. To avoid interrupting the flow of the presentation, all mathematical developments are relegated to the Appendix.

2. Dual divergences based estimates

The class of *dual* divergences estimators has been recently introduced by Keziou (2003) and Broniatowski and Keziou (2009). Recall that the ϕ-divergence between a bounded signed measure $\mathbb{Q}$, and a probability $\mathbb{P}$ on $\mathscr{D}$, when $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$, is defined by

$$D_\phi(\mathbb{Q},\mathbb{P}) := \int_{\mathscr{D}} \phi\left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)\, d\mathbb{P},$$

where ϕ is a convex function from $]-\infty,\infty[$ to $[0,\infty]$ with $\phi(1)=0$. We will consider only ϕ-divergences for which the function ϕ is strictly convex and satisfies: the domain of ϕ, $\mathrm{dom}\phi := \{x\in\mathbb{R} : \phi(x)<\infty\}$ is an interval with end points $a_\phi < 1 < b_\phi$, $\phi(a_\phi) = \lim_{x\downarrow a_\phi}\phi(x)$ and $\phi(a_\phi) = \lim_{x\uparrow b_\phi}\phi(x)$. The Kullback-Leibler, modified Kullback-Leibler, χ^2, modified χ^2 and Hellinger divergences are examples of ϕ-divergences; they are obtained respectively for $\phi(x) = x\log x - x + 1$, $\phi(x) = -\log x + x - 1$, $\phi(x) = \frac{1}{2}(x-1)^2$, $\phi(x) = \frac{1}{2}\frac{(x-1)^2}{x}$ and $\phi(x) = 2(\sqrt{x}-1)^2$. We extend the definition of these divergences on the whole space of all bounded signed measures via the extension of the definition of the corresponding ϕ functions on the whole real space $\mathbb{R}$ as follows: when ϕ is not well defined on $\mathbb{R}_-$ or well defined but not convex on $\mathbb{R}$, we set $\phi(x) = +\infty$ for all $x<0$. Observe for the χ^2-divergence, the corresponding ϕ function is defined on whole $\mathbb{R}$ and strictly convex. All the above examples are particular cases of the so-called "*power divergences*", introduced by Cressie and Read (1984) (see also Liese and Vajda (1987, Chapter 2)), which are defined through the class of convex real valued functions

$$x\in\mathbb{R}_+^* \to \phi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma-1)}$$

for γ in $\mathbb{R}\setminus\{0,1\}$, $\phi_0(x) := -\log x + x - 1$ and $\phi_1(x) := x\log x - x + 1$. (For all $\gamma\in\mathbb{R}$, we define $\phi_\gamma(0) := \lim_{x\downarrow 0}\phi_\gamma(x)$). So, the KL-divergence is associated to ϕ_1, the KL_m to ϕ_0, the χ^2 to ϕ_2, the χ^2_m to ϕ_{-1} and the Hellinger distance to $\phi_{1/2}$. In the monograph by Liese and Vajda (1987) the reader may find detailed ingredients of the modeling theory as well as surveys of the commonly used divergences.

Let $\{\mathbb{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta}\in\boldsymbol{\Theta}\}$ be some identifiable parametric model with $\boldsymbol{\Theta}$ a compact subset of $\mathbb{R}^d$. Consider the problem of estimation of the unknown true value of the parameter $\boldsymbol{\theta}_0$ on the basis of an i.i.d. sample $X_1,\ldots,X_n$. We shall assume that the observed data are from the probability space $(\mathcal{X}, \mathcal{A}, \mathbb{P}_{\boldsymbol{\theta}_0})$. Let ϕ be a function of class $\mathcal{C}^2$, strictly convex such that

$$\int \left| \phi' \left(\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(x)}{\mathrm{d}\mathbb{P}_{\boldsymbol{\alpha}}(x)} \right) \right| \, \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(x) < \infty, \forall \boldsymbol{\alpha} \in \boldsymbol{\Theta}. \tag{2.1}$$

Under assumption (2.1), using Fenchel duality technique, the divergence $D_\phi(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ can be represented as resulting from an optimization procedure, this result was elegantly proved in, Keziou (2003), Liese and Vajda (2006) and Broniatowski and Keziou (2009). Broniatowski and Keziou (2006) called it the *dual* form of a divergence, due to its connection with convex analysis. According to Liese and Vajda (2006), under the strict convexity and the differentiability of the function ϕ, it holds

$$\phi(t) \geq \phi(s) + \phi'(s)(t-s), \tag{2.2}$$

where the equality holds only for $s = t$. Let $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$ be fixed and put $t = \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(x)/\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_0}(x)$ and $s = \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(x)/\mathrm{d}\mathbb{P}_{\boldsymbol{\alpha}}(x)$ in (2.2) and then integrate with respect to $\mathbb{P}_{\boldsymbol{\theta}_0}$. This gives

$$D_\phi(\boldsymbol{\theta}, \boldsymbol{\theta}_0) := \int \phi \left(\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}}{\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_0}} \right) \, \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_0} = \sup_{\boldsymbol{\alpha} \in \boldsymbol{\Theta}} \int h(\boldsymbol{\theta}, \boldsymbol{\alpha}) \, \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}_0}, \tag{2.3}$$

where $h(\boldsymbol{\theta}, \boldsymbol{\alpha}, \cdot) : x \mapsto h(\boldsymbol{\theta}, \boldsymbol{\alpha}, x)$ and

$$h(\boldsymbol{\theta}, \boldsymbol{\alpha}, x) := \int \phi' \left(\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}}{\mathrm{d}\mathbb{P}_{\boldsymbol{\alpha}}} \right) \, \mathrm{d}\mathbb{P}_{\boldsymbol{\theta}} - \left[\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(x)}{\mathrm{d}\mathbb{P}_{\boldsymbol{\alpha}}(x)} \phi' \left(\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(x)}{\mathrm{d}\mathbb{P}_{\boldsymbol{\alpha}}(x)} \right) - \phi \left(\frac{\mathrm{d}\mathbb{P}_{\boldsymbol{\theta}}(x)}{\mathrm{d}\mathbb{P}_{\boldsymbol{\alpha}}(x)} \right) \right]. \tag{2.4}$$

Furthermore, the supremum in this display (2.3) is unique and reached in $\boldsymbol{\alpha} = \boldsymbol{\theta}_0$, independently upon the value of $\boldsymbol{\theta}$. Naturally, a class of estimators of $\boldsymbol{\theta}_0$, called "dual ϕ-divergence estimators" (DϕDE's), is defined by

$$\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) := \arg \sup_{\boldsymbol{\alpha} \in \boldsymbol{\Theta}} \mathbb{P}_n h(\boldsymbol{\theta}, \boldsymbol{\alpha}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \tag{2.5}$$

where $h(\boldsymbol{\theta}, \boldsymbol{\alpha})$ is the function defined in (2.4) and, for a measurable function f, $\mathbb{P}_n f$ denotes $n^{-1} \sum_{i=1}^n f(X_i)$. The class of estimators $\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})$ satisfies

$$\mathbb{P}_n \frac{\partial}{\partial \boldsymbol{\alpha}} h(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})) = 0. \tag{2.6}$$

Formula (2.5) defines a family of M-estimators indexed by the function ϕ specifying the divergence and by some instrumental value of the parameter $\boldsymbol{\theta}$.

In this section, we shall establish the consistency of bootstrapping under general conditions in the framework of dual divergence estimation. Define, for a measurable function f,

$$\mathbb{P}_n^* f := \frac{1}{n}\sum_{i=1}^{n} W_{ni} f(X_i),$$

where W_{ni}'s are the bootstrap weights defined on the probability space $(\mathcal{W}, \Omega, \mathbb{P}_W)$. In view of (2.5), the bootstrap estimator can be rewritten as

$$\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta}) := \arg\sup_{\boldsymbol{\alpha}\in\boldsymbol{\Theta}} \mathbb{P}_n^* h(\boldsymbol{\theta}, \boldsymbol{\alpha}). \tag{2.7}$$

The definition of $\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta})$, i.e., (2.7), implies that

$$\mathbb{P}_n^* \frac{\partial}{\partial \boldsymbol{\alpha}} h(\boldsymbol{\theta}, \widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta})) = 0. \tag{2.8}$$

The bootstrap weights W_{ni}'s are assumed to belong to the class of exchangeable bootstrap weights introduced in Præstgaard and Wellner (1993). To be precise, we shall assume the following conditions.

- W.1 *The vector $W_n = (W_{n1}, \ldots, W_{nn})'$ is exchangeable for all $n = 1, 2, \ldots$, i.e., for any permutation $\pi = (\pi_1, \ldots, \pi_n)$ of $(1, 2, \ldots, n)$, the joint distribution of $\pi(W_n) = (W_{n\pi_1}, \ldots, W_{n\pi_n})'$ is the same as that of W_n.*
- W.2 *$W_{ni} \geq 0$ for all n, i and $\sum_{i=1}^{n} W_{ni} = n$ for all n.*
- W.3 *$\limsup_{n\to\infty} \|W_{n1}\|_{2,1} \leq C < \infty$, where $\|W_{n1}\|_{2,1} = \int_0^\infty \sqrt{\mathbb{P}_W(W_{n1} \geq u)}du$.*
- W.4 *$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \sup_{t\geq\lambda} t^2 \mathbb{P}_W(W_{n1} > t) = 0$.*
- W.5 *$(1/n)\sum_{i=1}^{n}(W_{ni} - 1)^2 \xrightarrow{\mathbb{P}_W} c^2 > 0$.*

In Efron's nonparametric bootstrap, the bootstrap sample is drawn from the nonparametric estimate of the true distribution, i.e., empirical distribution. Thus, it is easy to show that $W_n \sim \text{Multinomial}(n; n^{-1}, \ldots, n^{-1})$ and conditions W.1–W.5 are satisfied. In general, conditions W.3-W.5 are easily satisfied under some moment conditions on W_{ni}, see Præstgaard and Wellner (1993, Lemma 3.1). In addition to Efron's nonparametric boostrap, the sampling schemes that satisfy conditions W.1–W.5 include *Bayesian bootstrap, Multiplier bootstrap, Double bootstrap, and Urn boostrap*. This list is sufficiently long to indicate that conditions W.1–W.5 are not unduely restrictive. Notice that the value of c in W.5 is independent of n and depends on the resampling method, e.g., $c = 1$ for the nonparametric bootstrap and Bayesian bootstrap, and $c = \sqrt{2}$ for the double bootstrap. A more precise discussion of this general formulation of the bootstrap can be found in Præstgaard and Wellner (1993), van der Vaart and Wellner (1996) and Kosorok (2008).

There exist two sources of randomness for the bootstrapped quantity, i.e., $\widehat{\boldsymbol{\alpha}}_{\phi}^{*}(\boldsymbol{\theta})$: the first comes from the observed data and the second is due to the resampling done by the bootstrap, i.e., random W_{ni}'s. Therefore, in order to rigorously state our main theoretical results for the general bootstrap of ϕ-divergence estimates, we need to specify relevant probability spaces and define stochastic orders with respect to relevant probability measures. We shall view X_i as the i-th coordinate projection from the canonical probability space $(\mathcal{X}^{\infty}, \mathcal{A}^{\infty}, \mathbb{P}_{\boldsymbol{\theta}_0}^{\infty})$ onto the i-th copy of $\mathcal{X}$. For the joint randomness involved, the product probability space is defined as

$$(\mathcal{X}^{\infty}, \mathcal{A}^{\infty}, \mathbb{P}_{\boldsymbol{\theta}_0}^{\infty}) \times (\mathcal{W}, \Omega, \mathbb{P}_W) = (\mathcal{X}^{\infty} \times \mathcal{W}, \mathcal{A}^{\infty} \times \Omega, \mathbb{P}_{\boldsymbol{\theta}_0}^{\infty} \times \mathbb{P}_W).$$

Throughout the paper, we assume that the bootstrap weights W_{ni}'s are independent of the data X_i's, thus

$$\mathbb{P}_{XW} = \mathbb{P}_{\boldsymbol{\theta}_0} \times \mathbb{P}_W.$$

Given a real-valued function Δ_n defined on the above product probability space, e.g. $\widehat{\boldsymbol{\alpha}}_{\phi}^{*}(\boldsymbol{\theta})$, we say that Δ_n is of an order $o_{\mathbb{P}_W}^{o}(1)$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability if for any $\epsilon, \eta > 0$,

$$\mathbb{P}_{\boldsymbol{\theta}_0}\{P_{W|X}^{o}(|\Delta_n| > \epsilon) > \eta\} \quad \longrightarrow \quad 0 \tag{2.9}$$

as $n \to 0$, and that Δ_n is of an order $O_{\mathbb{P}_W}^{o}(1)$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability if for any $\eta > 0$, there exists a $0 < M < \infty$ such that

$$\mathbb{P}_{\boldsymbol{\theta}_0}\{P_{W|X}^{o}(|\Delta_n| \geq M) > \eta\} \longrightarrow 0 \tag{2.10}$$

as $n \to \infty$, where the superscript "o" denotes the outer probability. For more details on the subject the interested reader may refer to Cheng and Huang (2010), in particular, Lemma 3 of the cited reference.

To establish the consistency of $\widehat{\boldsymbol{\alpha}}_{\phi}^{*}(\boldsymbol{\theta})$, the following conditions are assumed in our analysis.

(A.1)

$$\mathbb{P}_{\boldsymbol{\theta}_0} h(\boldsymbol{\theta}, \boldsymbol{\theta}_0) > \sup_{\boldsymbol{\alpha} \notin N(\boldsymbol{\theta}_0)} \mathbb{P}_{\boldsymbol{\theta}_0} h(\boldsymbol{\theta}, \boldsymbol{\alpha}) \tag{2.11}$$

for any open set $N(\boldsymbol{\theta}_0) \subset \boldsymbol{\Theta}$ containing $\boldsymbol{\theta}_0$.

(A.2)

$$\sup_{\boldsymbol{\alpha}\in\boldsymbol{\Theta}} |\mathbb{P}_n^* h(\boldsymbol{\theta},\boldsymbol{\alpha}) - \mathbb{P}_{\boldsymbol{\theta}_0} h(\boldsymbol{\theta},\boldsymbol{\alpha})| \xrightarrow{\mathbb{P}^o_{XW}} 0. \tag{2.12}$$

The following theorem gives the consistency of the bootstrapped estimates $\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta})$.

Theorem 2.1. *Assume that conditions* (A.1) *and* (A.2) *hold. Then* $\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta})$ *is a consistent estimate of* $\boldsymbol{\theta}_0$. *That is* $\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta}) \xrightarrow{\mathbb{P}^o_W} \boldsymbol{\theta}_0$ *in* $\mathbb{P}_{\boldsymbol{\theta}_0}$*-probability.*

The proof of Theorem 2.1 is postponed until §5.

Remark 2.1.
- *Condition* (A.1) *is the "well separated" condition, compactness of the parameter space* $\boldsymbol{\Theta}$ *and the continuity of divergence imply that the optimum is well-separated, provided the parametric model is identified, see van der Vaart (1998, Theorem 5.7).*
- *Condition* (A.2) *holds if the class* $\{h(\boldsymbol{\theta},\boldsymbol{\alpha}) : \boldsymbol{\alpha}\in\boldsymbol{\Theta}\}$ *is shown to be* P*-Glivenko-Cantelli, by applying van der Vaart and Wellner (1996, Lemma 3.6.16) and Cheng and Huang (2010, Lemma A.1).*

For any fixed $\delta_n > 0$, define the class of functions $\mathcal{H}_n$ and $\dot{\mathcal{H}}_n$ as

$$\mathcal{H}_n := \left\{ \frac{\partial}{\partial\boldsymbol{\alpha}} h(\boldsymbol{\theta},\boldsymbol{\alpha}) : \; \|\boldsymbol{\alpha}-\boldsymbol{\theta}_0\| \le \delta_n \right\} \tag{2.13}$$

and

$$\dot{\mathcal{H}}_n := \left\{ \frac{\partial^2}{\partial\boldsymbol{\alpha}^2} h(\boldsymbol{\theta},\boldsymbol{\alpha}) : \; \|\boldsymbol{\alpha}-\boldsymbol{\theta}_0\| \le \delta_n \right\}. \tag{2.14}$$

We shall say a class of functions $\mathcal{H} \in M(\mathbb{P}_{\boldsymbol{\theta}_0})$ if $\mathcal{H}$ possesses enough measurability for randomization with i.i.d. multipliers to be possible, i.e., $\mathbb{P}_n$ can be randomized, in other word, we can replace $(\delta_{X_i} - \mathbb{P}_{\boldsymbol{\theta}_0})$ by $(W_{ni}-1)\delta_{X_i}$. It is known that $\mathcal{H} \in M(\mathbb{P}_{\boldsymbol{\theta}_0})$, e.g., if $\mathcal{H}$ is countable, or if $\{\mathbb{P}_n\}_n^\infty$ are stochastically separable in $\mathcal{H}$, or if $\mathcal{H}$ is image admissible Suslin; see Giné and Zinn (1990, pages 853 and 854).
To state our result concerning the asymptotic normality, we shall assume the following additional conditions.

(A.3) The matrices $V := \mathbb{P}_{\boldsymbol{\theta}_0} \frac{\partial}{\partial\boldsymbol{\alpha}} h(\boldsymbol{\theta},\boldsymbol{\theta}_0) \frac{\partial}{\partial\boldsymbol{\alpha}} h(\boldsymbol{\theta},\boldsymbol{\theta}_0)^\top$ and $S := -\mathbb{P}_{\boldsymbol{\theta}_0} \frac{\partial^2}{\partial\boldsymbol{\alpha}^2} h(\boldsymbol{\theta},\boldsymbol{\theta}_0)$ are non singular.

(A.4) The class $\mathcal{H}_n \in M(\mathbb{P}_{\boldsymbol{\theta}_0}) \cap L_2(\mathbb{P}_{\boldsymbol{\theta}_0})$ and is $\mathbb{P}$-Donsker.

(A.5) The class $\dot{\mathcal{H}}_n \in M(\mathbb{P}_{\boldsymbol{\theta}_0}) \cap L_2(\mathbb{P}_{\boldsymbol{\theta}_0})$ and is $\mathbb{P}$-Donsker.

Conditions (A.4) and (A.5) ensure that the "size" of the function classes $\mathcal{H}_n$ and $\dot{\mathcal{H}}_n$ are reasonable so that the bootstrapped empirical processes $\mathbb{G}_n^* \equiv \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ indexed, respectively by $\mathcal{H}_n$ and $\dot{\mathcal{H}}_n$, have a limiting process conditional on the original observations; see Præstgaard and Wellner (1993, Theorem 2.2). The main result to be proved here may now be stated precisely as follows.

Theorem 2.2. *Assume that $\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})$ and $\widehat{\boldsymbol{\alpha}}^*_\phi(\boldsymbol{\theta})$ fullfil (2.6) and (2.8), respectively. In addition suppose that $\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} \boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\alpha}}^*_\phi(\boldsymbol{\theta}) \xrightarrow{\mathbb{P}^o_W} \boldsymbol{\theta}_0$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. Assume that conditions* (A.3–5) *and* W.1–W.5 *hold. Then we have*

$$\|\widehat{\boldsymbol{\alpha}}^*_\phi(\boldsymbol{\theta}) - \boldsymbol{\theta}_0\| = O^o_{\mathbb{P}_W}(n^{-1/2}) \tag{2.15}$$

in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. Furthermore,

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}^*_\phi(\boldsymbol{\theta}) - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})) = -S^{-1}\mathbb{G}_n^* \frac{\partial}{\partial \boldsymbol{\alpha}} h(\boldsymbol{\theta}, \boldsymbol{\theta}_0) + o^o_{\mathbb{P}_W}(1) \tag{2.16}$$

in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. Consequently,

$$\sup_{x\in\mathbb{R}^d} \left|\mathbb{P}_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\boldsymbol{\alpha}}^*_\phi(\boldsymbol{\theta}) - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})) \leq x) - \mathbb{P}(N(0,\Sigma) \leq x)\right| = o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1), \tag{2.17}$$

where " $\leq$ " is taken componentwise and c is given in W.5*, whose value depends on the used sampling scheme, and $\Sigma \equiv S^{-1}V(S^{-1})^\top$ where S and V are given in condition* (A.3)*. Thus, we have*

$$\sup_{x\in\mathbb{R}^d} \left|\mathbb{P}_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\boldsymbol{\alpha}}^*_\phi(\boldsymbol{\theta}) - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})) \leq x) - \mathbb{P}_{\boldsymbol{\theta}_0}(\sqrt{n}(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \boldsymbol{\theta}_0) \leq x)\right| \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} 0. \tag{2.18}$$

The proof of Theorem 2.1 is postponed until §5.

Remark 2.2. *Note that an appropriate choice of the the bootstrap weights W_{ni}'s implicates a smaller limit variance, that is, c^2 is smaller than 1. For instance, typical examples are i.i.d.-weighted bootstraps and the multivariate hypergeometric bootstrap, refer to Præstgaard and Wellner (1993, Examples 3.1 and 3.4).*

Following Cheng and Huang (2010), we shall illustrate how to apply our results to construct the confidence sets. A lower ϵ-th quantile of bootstrap distribution is defined to be any $q^*_{n\epsilon} \in \mathbb{R}^d$ fulfilling

$$q^*_{n\epsilon} := \inf\{x : \mathbb{P}_{W|\mathcal{X}_n}(\widehat{\boldsymbol{\alpha}}^*_\phi(\boldsymbol{\theta}) \leq x) \geq \epsilon\},$$

where x is an infimum over the given set only if there does not exist a $x_1 < x$ in $\mathbb{R}^d$ such that

$$\mathbb{P}_{W|\mathcal{X}_n}(\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta}) \leq x_1) \geq \epsilon.$$

Keep in mind the assumed regularity conditions on the criterion function, that is, $h(\boldsymbol{\theta}, \boldsymbol{\alpha})$ in the present framework, we can, without loss of generality, suppose that

$$\mathbb{P}_{W|\mathcal{X}_n}(\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta}) \leq q_{n\epsilon}^*) = \epsilon.$$

Making use the distribution consistency result given in (2.18), we can approximate the ϵ-th quantile of the distribution of $(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \boldsymbol{\theta}_0)$ by $(q_{n\epsilon}^* - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}))/c$. Therefore, we define the *percentile*-type bootstrap confidence set as

$$\mathrm{C}(\epsilon) := \left[\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) + \frac{q_{n(\epsilon/2)}^* - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})}{c}, \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) + \frac{q_{n(1-\epsilon/2)}^* - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})}{c}\right]. \tag{2.19}$$

In a similar manner, the ϵ-th quantile of $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \boldsymbol{\theta}_0)$ can be approximated by $\widetilde{q}_{n\epsilon}^*$, where $\widetilde{q}_{n\epsilon}^*$ is the ϵ-th quantile of the hybrid quantity $(\sqrt{n}/c)(\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta}) - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}))$, i.e.,

$$\mathbb{P}_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta}) - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})) \leq \widetilde{q}_{n\epsilon}^*) = \epsilon.$$

Note that $\widetilde{q}_{n\epsilon}^* = (\sqrt{n}/c)(q_{n\epsilon}^* - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}))$. Thus, the *hybrid*-type bootstrap confidence set would be defined as follows

$$\widetilde{\mathrm{C}}(\epsilon) := \left[\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \frac{\widetilde{q}_{n(1-\epsilon/2)}^*}{\sqrt{n}}, \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \frac{\widetilde{q}_{n(\epsilon/2)}^*}{\sqrt{n}}\right]. \tag{2.20}$$

Note that $q_{n\epsilon}^*$ and $\widetilde{q}_{n\epsilon}^*$ are not unique by the fact that we assume $\boldsymbol{\theta}$ is a vector. Recall that, for any $x \in \mathbb{R}^d$,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\theta}_0}(\sqrt{n}(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \boldsymbol{\theta}_0) \leq x) \quad &\longrightarrow \quad \Psi(x),\\
\mathbb{P}_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\boldsymbol{\alpha}}_\phi^*(\boldsymbol{\theta}) - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})) \leq x) \quad &\xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} \quad \Psi(x),
\end{aligned}$$

where $\Psi(x) = \mathbb{P}(N(0, \Sigma) \leq x)$. According to the quantile convergence Theorem, i.e., van der Vaart (1998, Lemma 21.1), we have, almost surely,

$$\widetilde{q}_{n\epsilon}^* \xrightarrow{\mathbb{P}_{XW}} \Psi^{-1}(\epsilon).$$

When applying quantile convergence theorem, we use the almost sure representation, that is, van der Vaart (1998, Theorem 2.19), and argue along subsequences.

Considering the Slutsky's Theorem which ensures that $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \boldsymbol{\theta}_0) - \widetilde{q}^*_{n(\epsilon/2)}$ weakly converges to $N(0, \Sigma) - \Psi^{-1}(\epsilon/2)$, we further have

$$\begin{aligned}\mathbb{P}_{XW}\left(\boldsymbol{\theta}_0 \leq \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \frac{\widetilde{q}^*_{n(\epsilon/2)}}{\sqrt{n}}\right) &= \mathbb{P}_{XW}\left(\sqrt{n}(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \boldsymbol{\theta}_0) \geq \widetilde{q}^*_{n(\epsilon/2)}\right) \\ &\longrightarrow \mathbb{P}_{XW}\left(N(0, \Sigma) \geq \Psi^{-1}(\epsilon/2)\right) = 1 - \epsilon/2.\end{aligned}$$

The above arguments prove the consistency of the *hybrid*-type bootstrap confidence set, i.e., (2.22), and can also be applied to the *percentile*-type bootstrap confidence set, i.e., (2.21). For an in-depth study and more rigorous proof, we may refer to van der Vaart (1998, Lemma 23.3). The above discussion may be summarized as follows.

Corollary 2.3. *Under the conditions in Theorem 2.2, we have, as $n \to \infty$,*

$$\mathbb{P}_{XW}\left(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) + \frac{q^*_{n(\epsilon/2)} - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})}{c} \leq \boldsymbol{\theta}_0 \leq \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) + \frac{q^*_{n(1-\epsilon/2)} - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})}{c}\right) \longrightarrow 1 - \epsilon, \tag{2.21}$$

$$\mathbb{P}_{XW}\left(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \frac{\widetilde{q}^*_{n(1-\epsilon/2)}}{\sqrt{n}} \leq \boldsymbol{\theta}_0 \leq \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}) - \frac{\widetilde{q}^*_{n(\epsilon/2)}}{\sqrt{n}}\right) \longrightarrow 1 - \epsilon. \tag{2.22}$$

It is well known that the above bootstrap confidence sets can be obtained easily through routine bootstrap sampling.

Remark 2.3. *Notice that the choice of weights depends on the problem at hand : accuracy of the the estimation of the entire distribution of the statistic, accuracy of a confidence interval, accuracy in large deviation sense, accuracy for a finite sample size. Barbe and Bertail (1995) indicate that the area where the weighted bootstrap clearly performs better than the classical bootstrap is in term of coverage accuracy.*

3. Random right censoring

Let $T_1, \ldots, T_n$ be i.i.d. survival times with continuous survival function $\overline{F}(\cdot) = 1 - F(\cdot)$ and $C_1, \ldots, C_n$ be independent censoring times with d.f. $G(\cdot)$. In the censoring set-up, we observe only the pair $Y_i = \min(T_i, C_i)$ and $\delta_i = \mathbb{1}\{T_i \leq C_i\}$, which designs whether an observation has been censored or not. Let $(Y_1, \delta_1), \ldots, (Y_n, \delta_n)$ denote the observed data points and $t(1) < t(2) < \cdots < t(k)$ be the k distinct death times. Now define the death set and risk set as follows for $j = 1, \ldots, k$:

$$D(j) := \{i : y_i = t(j), \delta_i = 1\}$$

and

$$R(j) := \{i : y_i \geq t(j)\}.$$

The Kaplan and Meier (1958)'s estimator of $\overline{F}(\cdot)$ may be written as follows

$$\overline{F}_n(t) := \prod_{j=1}^{k} \left(1 - \frac{\sum_{q\in D(j)} 1}{\sum_{q\in R(j)} 1}\right)^{\mathbb{1}\{T_{(j)}\leq t\}}.$$

One may define a generally exchangeable weighted bootstrap scheme for the Kaplan-Meier estimator and related functionals as follows, cf. James (1997, p. 1598),

$$1 - F_n^*(t) := \overline{F}_n^*(t) = \prod_{j=1}^{k} \left(1 - \frac{\sum_{q\in D(j)} W_{nq}}{\sum_{q\in R(j)} W_{nq}}\right)^{\mathbb{1}\{T(j)\leq t\}}.$$

Let ψ be F-integrable and put

$$\Psi_n := \int \psi(u) dF_n^*(u) = \sum_{j=1}^{k} \Upsilon_{jn}\psi(T_{(j)}),$$

where

$$\Upsilon_{jn} := \left(\frac{\sum_{q\in D(j)} W_{nq}}{\sum_{q\in R(j)} W_{nq}}\right) \prod_{k=1}^{j-1} \left(\frac{\sum_{q\in D(k)} W_{nq}}{\sum_{q\in R(k)} W_{nq}}\right).$$

Note that we have used the following fact. Let a_i, $i = 1, \ldots, k$, b_i, $i = 1, \ldots, k$, be real numbers

$$\prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i = \sum_{i=1}^{k} (a_i - b_i) \prod_{j=1}^{i-1} b_j \prod_{h=1+i}^{k} a_h.$$

In the similar way, we define a more appropriate representation, that will be used in the sequel, as follows

$$\Psi_n = \int \psi(u) dF_n^*(u) = \sum_{j=1}^{n} \pi_{jn}\psi(Y_{j:n}),$$

where

$$\pi_{jn} := \delta_{j:n} \left(\frac{\sum_{q\in D(j)} W_{nq}}{\sum_{q\in R(j)} W_{nq}}\right) \prod_{k=1}^{j-1} \left(\frac{\sum_{q\in D(k)} W_{nq}}{\sum_{q\in R(k)} W_{nq}}\right)^{\delta_{k:n}}.$$

Here, $Y_{1:n} \leq \cdots \leq Y_{n:n}$ are ordered Y-values and $\delta_{i:n}$ denotes the concomitant associated with $Y_{i:n}$. For the right censoring situation, the bootstrap "dual ϕ-divergence estimators" (DϕE's), is defined by replacing $\mathbb{P}_n$ in (2.5) by $\widehat{P}_n^*$, that is

$$\widehat{\boldsymbol{\alpha}}_n(\boldsymbol{\theta}) := \arg\sup_{\boldsymbol{\alpha}\in\boldsymbol{\Theta}} \int h(\boldsymbol{\theta}, \boldsymbol{\alpha}) \mathrm{d}\widehat{P}_n^*, \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}. \tag{3.1}$$

The corresponding estimating equation for the unknown parameter is then given by

$$\int \frac{\partial}{\partial \boldsymbol{\alpha}} h(\boldsymbol{\theta}, \boldsymbol{\alpha}) \mathrm{d}\widehat{P}_n^* = 0. \tag{3.2}$$

Formula (3.1) defines a family of M-estimator for censored data. For more details about dual ϕ-divergence estimators in right censoring we refer to Cherfi (2011), we leave this study open for future research.

4. Simulations

In this section, series of experiments were conducted in order to examine the performance of the proposed random weighted bootstrap procedure of the dual ϕ-divergence estimators (DϕDE), defined in (2.7). We provide numerical illustrations regarding the mean squared error and the coverage probabilities. The computing program codes were implemented in `R`.

The values of γ are chosen to be $-1,\ 0,\ 0.5,\ 1,\ 2$, which corresponds, as indicated above, to the well known standard divergences: χ^2_m-divergence, KL_m, the Hellinger distance, KL and the χ^2-divergence respectively. The samples of sizes considered in our simulations are $25,\ 50,\ 75,\ 100,\ 150,\ 200$ and the estimates, DϕDE's $\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})$, are obtained from 500 independent runs. The value of escort parameter θ is taken to be the MLE, which under the model is a consistent estimate of θ_0, and the limit distribution of the DϕDE $\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}_0)$, in this case, has variance which indeed coincides with the MLE, for more details on this subject, we refer to Keziou (2003, Theorem 2.2, (1) (b)). The bootstrap weights are chosen $(W_{n1}, \ldots, W_{nn}) \sim$ Dirichlet$(n; 1, \ldots, 1)$.

In Figure 1, we plot the densities of the different estimates, it shows that the proposed estimators perform reasonably well.

Table 2 provides the MSE of various estimates under the Normal model $N(\theta_0 = 0, 1)$. Here, we mention that the KL based estimator ($\gamma = 1$) is more efficient than the others competitors.

Table 4 provides the MSE of various estimates under the Exponential model $\exp(\theta_0 = 1)$. As expected, the MLE produces most efficient estimators in the . A close look at the results of the simulations show that the DϕDE's performs well under the model. For large sample size $n = 200$, the estimator based on the Hellinger distance is equivalent to that of the MLE. Indeed in terms of empirical MSE the DϕDE's

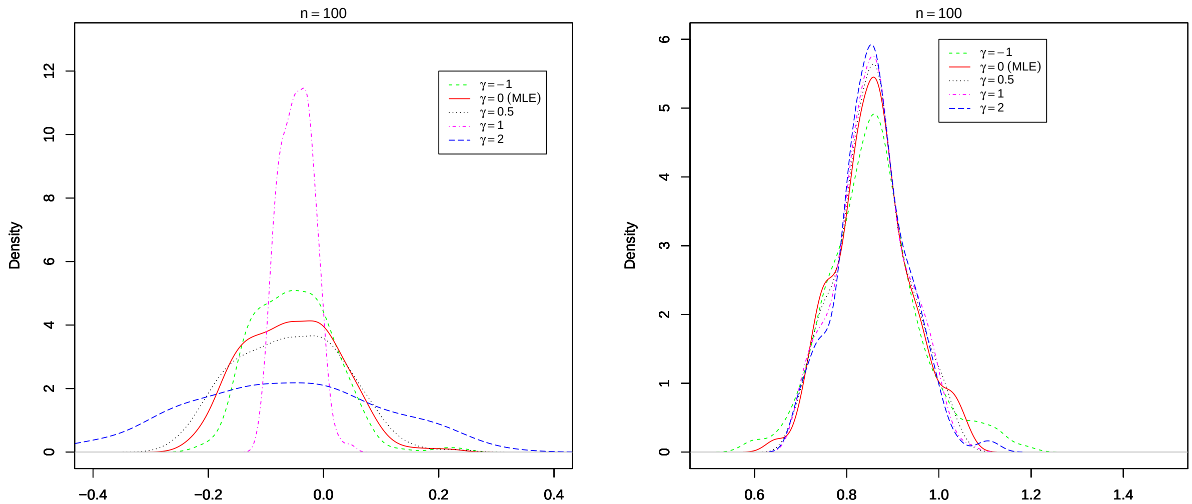

FIGURE 1. Densities of the estimates.

with $\gamma = 0.5$ produces the same MSE as the MLE, while the performance of the other estimators is comparable.

TABLE 1. MSE of the estimates for the Normal distribution, B=500

	$n = 25$	$n = 50$	$n = 75$	$n = 100$	$n = 150$	$n = 200$
1	0.0687	0.0419	0.0288	0.0210	0.0135	0.0107
2	0.0647	0.0373	0.0255	0.0192	0.0127	0.0101
3	0.0668	0.0379	0.0257	0.0194	0.0128	0.0101
4	0.0419	0.0217	0.0143	0.0108	0.0070	0.0057
5	0.0931	0.0514	0.0331	0.0238	0.0148	0.0112

Tables 6 and 8, provide the empirical coverage probabilities of the corresponding 0.95 weighted bootstrap confidence intervals based on $B = 500, 1000$ weighted bootstrap estimators. Notice that the empirical coverage probabilities as in any other inferential context, the greater the sample size, the better. From the results reported in these tables, we find that for large values of the sample size n, the the empirical coverage probabilities are all close to the nominal level. One can see that

Table 2. MSE of the estimates for the Normal distribution, B=1000

γ	$n=25$	$n=50$	$n=75$	$n=100$	$n=150$	$n=200$
-1	0.0716	0.0432	0.0285	0.0224	0.0147	0.0099
0	0.0670	0.0385	0.0255	0.0202	0.0136	0.0093
0.5	0.0684	0.0391	0.0258	0.0203	0.0137	0.0093
1	0.0441	0.0230	0.0143	0.0116	0.0078	0.0049
2	0.0900	0.0522	0.0335	0.0246	0.0156	0.0103

Table 3. MSE of the estimates for the Exponential distribution, B=500

	$n=25$	$n=50$	$n=75$	$n=100$	$n=150$	$n=200$
1	0.0729	0.0435	0.0313	0.0215	0.0146	0.0117
2	0.0708	0.0405	0.0280	0.0195	0.0131	0.0104
3	0.0727	0.0415	0.0282	0.0197	0.0131	0.0105
4	0.0786	0.0446	0.0296	0.0207	0.0136	0.0108
5	0.1109	0.0664	0.0424	0.0289	0.0178	0.0132

Table 4. MSE of the estimates for the Exponential distribution, B=1000

γ	$n=25$	$n=50$	$n=75$	$n=100$	$n=150$	$n=200$
-1	0.0670	0.0444	0.0295	0.0243	0.0146	0.0111
0	0.0659	0.0417	0.0269	0.0216	0.0133	0.0102
0.5	0.0677	0.0427	0.0272	0.0216	0.0135	0.0102
1	0.0735	0.0458	0.0287	0.0225	0.0140	0.0106
2	0.1074	0.0697	0.0429	0.0306	0.0183	0.0133

the DϕDE's with $\gamma = 2$ have the best empirical coverage probability which is near the assigned nominal level.

5. Appendix

This section is devoted to the proofs of our results.

TABLE 5. Empirical coverage probabilities for the Normal distribution, B=500

	$n=25$	$n=50$	$n=75$	$n=100$	$n=150$	$n=200$
1	0.88	0.91	0.93	0.92	0.95	0.92
2	0.91	0.92	0.94	0.94	0.94	0.93
3	0.94	0.94	0.94	0.96	0.94	0.93
4	0.44	0.47	0.54	0.46	0.48	0.51
5	0.97	0.97	0.96	0.97	0.95	0.95

TABLE 6. Empirical coverage probabilities for the Normal distribution, B=1000

	$n=25$	$n=50$	$n=75$	$n=100$	$n=150$	$n=200$
γ						
-1	0.87	0.90	0.93	0.92	0.93	0.96
0	0.91	0.94	0.94	0.93	0.94	0.96
0.5	0.93	0.93	0.95	0.93	0.94	0.96
1	0.46	0.45	0.48	0.46	0.45	0.50
2	0.96	0.97	0.96	0.95	0.96	0.96

TABLE 7. Empirical coverage probabilities for the Exponential distribution, B=500

	$n=25$	$n=50$	$n=75$	$n=100$	$n=150$	$n=200$
1	0.67	0.83	0.87	0.91	0.93	0.92
2	0.73	0.87	0.91	0.93	0.96	0.93
3	0.76	0.88	0.91	0.94	0.96	0.93
4	0.76	0.88	0.90	0.95	0.97	0.93
5	0.76	0.89	0.91	0.96	0.96	0.94

5.1. **Proof of Theorem 2.1.** Proceeding as van der Vaart and Wellner (1996) in their proof of Corollary 3.2.3, it is straightforward to show the consistency of $\widehat{\boldsymbol{\alpha}}_{\phi}^{*}(\boldsymbol{\theta})$.

$\square$

TABLE 8. Empirical coverage probabilities for the Exponential distribution, B=1000

	$n=25$	$n=50$	$n=75$	$n=100$	$n=150$	$n=200$
γ						
-1	0.70	0.79	0.90	0.91	0.92	0.91
0	0.76	0.84	0.91	0.92	0.93	0.92
0.5	0.78	0.85	0.93	0.94	0.94	0.93
1	0.78	0.87	0.94	0.94	0.95	0.94
2	0.78	0.88	0.95	0.95	0.96	0.95

Remark 5.1. *Note that the proof techniques of Theorem 2.2 are largely inspired from that of Cheng and Huang (2010) and changes have been made in order to adapt them to our purpose.*

5.2. **Proof of Theorem 2.2.** Keep in mind the following definitions

$$\mathbb{G}_n := \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{\boldsymbol{\theta}_0})$$

and

$$\mathbb{G}_n^* := \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n).$$

In view of the fact that $\mathbb{P}_{\boldsymbol{\theta}_0}\dfrac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) = 0$, then a little calculation shows that

$$\begin{aligned}
&\mathbb{G}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) + \mathbb{G}_n\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) + \sqrt{n}\mathbb{P}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right] \\
&= \mathbb{G}_n^*\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)\right] + \mathbb{G}_n\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)\right] \\
&\quad + \sqrt{n}\mathbb{P}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*).
\end{aligned}$$

Consequently, we have following inequality

$$\begin{aligned}
&\left\|\sqrt{n}\mathbb{P}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)-\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right]\right\| \\
\leq\ & \left\|\mathbb{G}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right\|+\left\|\mathbb{G}_n\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right\| \\
&+\left\|\mathbb{G}_n^*\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)-\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right]\right\| \\
&+\left\|\mathbb{G}_n\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)-\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right]\right\| \\
&+\left\|\sqrt{n}\mathbb{P}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)\right\| \\
:=\ & G_1+G_2+G_3+G_4+G_5. \qquad (5.1)
\end{aligned}$$

According to Theorem 2.2 in Præstgaard and Wellner (1993), under condition (A.4), we have $G_1=O^o_{\mathbb{P}_W}(1)$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. In view of the CLT, we have $G_2=O_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1)$. By applying a Taylor series expansion, we have

$$\mathbb{G}_n^*\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)-\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right]=(\widehat{\boldsymbol{\alpha}}^*-\boldsymbol{\theta}_0)^\top\,\mathbb{G}_n^*\frac{\partial^2}{\partial\boldsymbol{\alpha}^2}h(\boldsymbol{\theta},\overline{\boldsymbol{\alpha}}), \qquad (5.2)$$

where $\overline{\boldsymbol{\alpha}}$ is between $\widehat{\boldsymbol{\alpha}}^*$ and $\boldsymbol{\theta}_0$. By condition (A.5) and Theorem 2.2 in Præstgaard and Wellner (1993), we conclude that the right term in (5.2) is of order $O^o_{\mathbb{P}_W}(\|\widehat{\boldsymbol{\alpha}}^*-\boldsymbol{\theta}_0\|)$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. The fact that $\widehat{\boldsymbol{\alpha}}^*$ is assumed to be consistent, then, we have $G_3=o^o_{\mathbb{P}_W}(1)$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. An analogous argument yields

$$\mathbb{G}_n\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)-\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right]$$

is of order $O_{\mathbb{P}_{\boldsymbol{\theta}_0}}(\|\widehat{\boldsymbol{\alpha}}^*-\boldsymbol{\theta}_0\|)$, by the consistency of $\widehat{\boldsymbol{\alpha}}^*$, we have $G_4=o^o_{\mathbb{P}_W}(1)$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. Finally, $G_5=0$ based on (2.8). In summary, (5.1) can be rewritten as follows

$$\left\|\sqrt{n}\mathbb{P}_{\boldsymbol{\theta}_0}(\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*)-\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0))\right\|\leq O^o_{\mathbb{P}_W}(1)+O^o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1) \qquad (5.3)$$

in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. On the other hand, by a Taylor series expansion, we can write

$$\mathbb{P}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\alpha})-\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right]=-(\boldsymbol{\alpha}-\boldsymbol{\theta}_0)^\top S+O\left(\|\boldsymbol{\alpha}-\boldsymbol{\theta}_0\|^2\right). \qquad (5.4)$$

Clearly it is straightforward to combine (5.4) with (5.3) to infer the following

$$\sqrt{n}\,\|S\|\widehat{\boldsymbol{\alpha}}^* - \boldsymbol{\theta}_0\|\| \leq O^o_{\mathbb{P}_W}(1) + O^o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1) + O^o_{\mathbb{P}_W}\left(\sqrt{n}\|\widehat{\boldsymbol{\alpha}}^* - \boldsymbol{\theta}_0\|^2\right) \tag{5.5}$$

in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. By considering again the consistency of $\widehat{\boldsymbol{\alpha}}^*$ and condition (A.3) and making use (5.5) to complete the proof of (2.15).

We next prove (2.16). Introduce

$$\begin{aligned}
H_1 &:= -\mathbb{G}_n^*\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right],\\
H_2 &:= \mathbb{G}_n\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right],\\
H_3 &:= -\mathbb{G}_n\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right],\\
H_4 &:= \sqrt{n}\mathbb{P}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*) - \sqrt{n}\mathbb{P}_n\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}).
\end{aligned}$$

By some algebra, we obtain

$$\sqrt{n}\mathbb{P}_{\boldsymbol{\theta}_0}\left(\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}})\right) + \mathbb{G}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) = \sum_{j=1}^4 H_j.$$

Obviously, $H_1 = O^o_{\mathbb{P}_W}(n^{-1/2})$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability and $H_2 = O_{\mathbb{P}_{\boldsymbol{\theta}_0}}(n^{-1/2})$. We also know that the order of H_3 is $O^o_{\mathbb{P}_W}(n^{-1/2})$ in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. Using (2.6) and (2.8) we obtain that $H_4 = 0$.

Therefore, we have established

$$\sqrt{n}\mathbb{P}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}})\right] = -\mathbb{G}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) + o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1) + o^o_{\mathbb{P}_W}(1) \tag{5.6}$$

in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. To analyze the left hand side of (5.6), we rewrite it as

$$\sqrt{n}\mathbb{P}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}^*) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right] - \sqrt{n}\mathbb{P}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\widehat{\boldsymbol{\alpha}}) - \frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0)\right].$$

By a Taylor expansion, we obtain

$$\begin{aligned}
&\sqrt{n}S(\widehat{\boldsymbol{\alpha}}^*_\phi(\boldsymbol{\theta}) - \widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta}))\\
&= \mathbb{G}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) + o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1) + o^o_{\mathbb{P}_W}(1) + O_{\mathbb{P}_{\boldsymbol{\theta}_0}}(n^{-1/2}) + O^o_{\mathbb{P}_W}(n^{-1/2})\\
&= \mathbb{G}_n^*\frac{\partial}{\partial\boldsymbol{\alpha}}h(\boldsymbol{\theta},\boldsymbol{\theta}_0) + o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1) + o^o_{\mathbb{P}_W}(1)
\end{aligned} \tag{5.7}$$

in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability. Keep in mind that, under condition (A.3), the matrix S is non-singular. Multiply both sides of (5.7) by S^{-1} to obtain (2.16). An application of Præstgaard and Wellner (1993, Lemma 4.6), under the bootstrap weight conditions, thus implies (2.17). Using Broniatowski and Keziou (2009, Theorem 3.2) and van der Vaart (1998, Lemma 2.11), it easily follows that

$$\sup_{x\in\mathbb{R}^d}\left|\mathbb{P}_{\boldsymbol{\theta}_0}(\sqrt{n}(\widehat{\boldsymbol{\alpha}}_\phi(\boldsymbol{\theta})-\boldsymbol{\theta}_0)\leq x)-\mathbb{P}(N(0,\Sigma)\leq x)\right| = o_{\mathbb{P}_{\boldsymbol{\theta}_0}}(1). \tag{5.8}$$

By combining (2.17) and (5.8), we readily obtain the desired conclusion (2.18).

□

References

Barbe, P. and Bertail, P. (1995). *The weighted bootstrap*, volume 98 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Beran, R. (1984). Bootstrap methods in statistics. *Jahresber. Deutsch. Math.-Verein.*, **86**(1), 14–30.

Beran, R. and Millar, P. W. (1986). Confidence sets for a multivariate distribution. *Ann. Statist.*, **14**(2), 431–443.

Beran, R. J., Le Cam, L., and Millar, P. W. (1987). Convergence of stochastic empirical measures. *J. Multivariate Anal.*, **23**(1), 159–168.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, **9**(6), 1196–1217.

Broniatowski, M. and Keziou, A. (2006). Minimization of ϕ-divergences on sets of signed measures. *Studia Sci. Math. Hungar.*, **43**(4), 403–442.

Broniatowski, M. and Keziou, A. (2009). Parametric estimation and tests through divergences and the duality technique. *J. Multivariate Anal.*, **100**(1), 16–36.

Chatterjee, S. and Bose, A. (2005). Generalized bootstrap for estimating equations. *Ann. Statist.*, **33**(1), 414–436.

Cheng, G. and Huang, J. (2010). Bootstrap consistency for general semiparametric M-estimation *Ann. Statist.*, **38**(5), 2884–2915.

Cherfi, M. (2011). Dual divergences estimation for censored survival data. *Submitted.*

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B*, **46**(3), 440–464.

Csörgő, S. and Mason, D. M. (1989). Bootstrapping empirical functions. *Ann. Statist.*, **17**(4), 1447–1471.

del Barrio, E. and Matrán, C. (2000). The weighted bootstrap mean for heavy-tailed distributions. *J. Theoret. Probab.*, **13**(2), 547–569.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, **7**(1), 1–26.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, volume 57 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York.

Gaenssler, P. (1992). Confidence bands for probability distributions on Vapnik-Chervonenkis classes of sets in arbitrary sample spaces using the bootstrap. In *Bootstrapping and related techniques (Trier, 1990)*, volume 376 of *Lecture Notes in Econom. and Math. Systems*, pages 57–61. Springer, Berlin.

Giné, E. and Zinn, J. (1989). Necessary conditions for the bootstrap of the mean. *Ann. Statist.*, **17**(2), 684–691.

Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.*, **18**(2), 851–869.

Holmes, S. and Reinert, G. (2004). Stein's method for the bootstrap. In *Stein's method: expository lectures and applications*, volume 46 of *IMS Lecture Notes Monogr. Ser.*, pages 95–136. Inst. Math. Statist., Beachwood, OH.

Keziou, A. (2003). Dual representation of ϕ-divergences and applications. *C. R. Math. Acad. Sci. Paris*, **336**(10), 857–862.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Liese, F. and Vajda, I. (1987). *Convex statistical distances*, volume 95 of *Teubner-Texte zur Mathematik [Teubner Texts in Mathematics]*. BSB B. G. Teubner Verlagsgesellschaft, Leipzig. With German, French and Russian summaries.

Liese, F. and Vajda, I. (2006). On divergences and informations in statistics and information theory. *IEEE Trans. Inform. Theory*, **52**(10), 4394–4412.

Lo, A. Y. (1993). A Bayesian bootstrap for censored data. *Ann. Statist.*, **21**(1), 100–123.

Lohse, K. (1987). Consistency of the bootstrap. *Statist. Decisions*, **5**(3-4), 353–366.

James, L. F. (1997). A study of a class of weighted bootstraps for censored data. *Ann. Statist.*, **25**(4), 1595–1621.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457–481.

Mason, D. M. and Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.*, **20**(3), 1611–1624.

Pardo, L. (2006). *Statistical inference based on divergence measures*, volume 185 of *Statistics: Textbooks and Monographs*. Chapman & Hall/CRC, Boca Raton, FL.

Præstgaard, J. and Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.*, **21**(4), 2053–2086.

Rubin, D. B. (1981). The Bayesian bootstrap. *Ann. Statist.*, **9**(1), 130–134.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Weng, C.-S. (1989). On a second-order asymptotic property of the Bayesian bootstrap mean. *Ann. Statist.*, **17**(2), 705–710.

Zheng, Z. G. and Tu, D. S. (1988). Random weighting methods in regression models. *Sci. Sinica Ser. A*, **31**(12), 1442–1459.

L.S.T.A., Université Pierre et Marie Curie, 4 place Jussieu 75252 Paris Cedex 05, France; e-mail adresses : salim.bouzebda@upmc.fr ; mohamed.cherfi@gmail.com