Theory of Stochastic Processes
Vol. 25 (41), no. 1, 2020, pp. 1–24

SALIM BOUZEBDA AND MOHAMED CHERFI

GENERAL INFERENCE IN SEMIPARAMETRIC MODELS THROUGH DIVERGENCES AND THE DUALITY TECHNIQUE WITH APPLICATIONS

In this paper, we extend the dual divergence approach to general semiparametric models and study dual divergence estimators for semiparametric models. Asymptotic properties such as consistency, asymptotic normality of the proposed estimators are deeply investigated by mean the sophisticated modern empirical theory. We investigate the exchangeably weighted estimators in this setting and establish the consistency. We finally consider the functional M-estimator and obtain its weak convergence result.

1. Introduction

The ϕ-divergence modeling has proved to be a flexible tool and provided a powerful statistical modeling framework in a variety of applied and theoretical contexts [refer to [15], [42] and [33, 32] and the references therein]. For good recent sources of references to research literature in this area along with statistical applications consult [5] and [42]. The main aim in the parametric estimation are efficiency when the model has been appropriately chosen and to attain robustness (against model misspecification) when it has not. Notice that the major practical problem of maximum likelihood estimators is the lack of robustness, while many robust estimators achieve robustness at some cost in first-order efficiency. The appeal of dual divergences method is that in addition to the statistical efficiency of the estimators when the parametric model is correctly specified, these estimators are also robust to contamination, for a deep investigation regarding this issue we may refer to [45]. Efficiency combined with robustness properties dual divergences estimators appealing in practice and form a desirable class of estimators. In general, [45] proved that the dual divergence estimators have excellent robustness properties for parametric models, such as resistance to outliers as well as robustness with respect to model misspecification. Furthermore, an appropriate choice of divergence may of special attraction that it is dimensionless, such is the case for Hellinger divergence. Application of dual representation of ϕ-divergences have been considered by many authors, we cite among others, [30] for semi-parametric two-sample density ratio models, robust tests based on saddlepoint approximations in [46], [45] have proved that this class contains robust and efficient estimators and proposed robust test statistics based on divergences estimators. An extension of dual ϕ-divergences estimators to right censored data are introduced in [23], for estimation and tests in copula models we refer to [11] and the references therein. Performances of dual ϕ-divergence estimators for normal models are studied in [22]. Unfortunately in the preceding paper, 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, [10] propose a general bootstrap of ϕ-divergence based estimators. Throughout the available literature,

2000 *Mathematics Subject Classification.* Primary: 62F40; 62F35; 62F12; 62G20; 62G09; Secondary: 62G30.

Key words and phrases. Divergences; M-estimators; Robust estimation; Semiparametric; Minimum distance estimators; empirical processes.

investigations on the asymptotic properties of dual divergence estimators, as well as the relevant test statistics, have privileged the parametric case. However, in practice, we need more flexible models that contain both parametric and nonparametric components with are the semiparamtric models, i.e., statistical models where at least one parameter of interest is not Euclidean. This paper concentrates on this specific problem. We aim, namely, to investigate semiparametric inference procedures by introducing a semiparametric inference procedure based on *divergence* and *duality* technique and derive general theorems on the asymptotic behavior of the proposed estimators. To the best of our knowledge, the results, presented in this paper, are believed to be novel and the problem of semi parametric inference by the mean of dual divergence have not been tackled in the literature.

The layout of the present article is structured as follows. Section 2 is devoted to the definitions and notations needed to state our main results as well as the estimators that we are interested in. In Section 3, give our main theoretical results including the consistency and the asymptotic distribution of the semiparametric ϕ-divergence estimate. Some examples are given in Section 4. In Section 5, we investigate the bootstrapped estimators. We consider a class of functional M-estimator processes in Section 6. To prevent from interrupting the flow of the presentation, all proofs are gathered in Section 8.

2. Model and estimation procedures

2.1. Model and mathematical backgrounds

To formulate the problem that we will treat in this paper, we need the following notation. Let μ be fixed σ-finite measure on $(\mathcal{X}, \mathcal{A})$, where $\mathcal{X}$ is the sample space and $\mathcal{A}$ is a σ-algebra of subsets of $\mathcal{X}$. Suppose that the unknown probability measure $\mathbb{P}_{\theta,\eta}$ on $(\mathcal{X}, \mathcal{A})$ is dominated by μ. In our framework, θ is the parameter of primary interest while η is needed only to describe the model. Consider a set of independent random variables $X_1, \ldots, X_n$ to be observed from probability density function $\mathbb{P}_{\theta,\eta}$. In the sequel, we assume that the common density $\mathbb{P}_{\theta,\eta}$ is a member of the semiparametric model

$$\mathcal{P} := \{\mathbb{P}_{\theta,\eta} : \mathbb{P}_{\theta,\eta} := d\mathbb{P}_{\theta,\eta}/d\mu, \theta \in \Theta, \eta \in \mathcal{H}\}, \tag{2.1}$$

where θ is a Euclidean parameter in $\Theta \in \mathbb{R}^p$ and η belongs to an infinite-dimensional set $\mathcal{H}$. In our framework, θ, commonly called Euclidean parameter, is the parameter of primary interest while η is a nuisance parameter needed only to describe the model. Throughout this paper, the true value $(\theta, \eta) \in \Theta \times \mathcal{H}$ is denoted by (θ_0, η_0). Numerous examples fall into the class (2.1), well-known examples include semiparametric mixture models [47], errors-in-variables models [6] and [39], regression models [50]. More examples and theory can be found in the monographs of [7], [50], [48] and [31] and in the references therein. First, we shall introduce some notation and definitions which will be used for the statement of our forthcoming results. Recall that the ϕ-divergence between a bounded signed measure $\mathbb{Q}$, and a probability measure $\mathbb{P}$ on $\mathscr{D}$, when $\mathbb{Q}$ is absolutely continuous with respect to $\mathbb{P}$, is defined by

$$D_\phi(\mathbb{Q}, \mathbb{P}) := \int_{\mathcal{X}} \phi\left(\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right) \, \mathrm{d}\mathbb{P},$$

where $\phi(\cdot)$ 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 $\phi(x) = 2(\sqrt{x}-1)^2$. All these divergences, belong to the class of the so called "power divergences" introduced in [24] (see also [32] chapter 2). They are defined through the class of convex functions

$$x \in]0,+\infty[\mapsto \phi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma-1)}, \tag{2.2}$$

if $\gamma \in \mathbb{R} \setminus \{0,1\}$, $\phi_0(x) := -\log x + x - 1$ 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 $\phi_{1/2}$. We refer to [32] for an overview on the origin of the concept of divergences in statistics.
Let ϕ be a function of class $\mathcal{C}^2$, strictly convex and satisfies

$$\int \left| \phi'\left(\frac{\mathbb{P}_{\theta,\eta}(x)}{\mathbb{P}_{\alpha,\eta}(x)} \right) \right| \, \mathrm{d}\mathbb{P}_{\theta,\eta}(x) < \infty. \tag{2.3}$$

As it is mentioned in [15], if the function $\phi(\cdot)$ satisfies the following conditions

$$\begin{array}{c} \text{there exists } \ 0 < \delta < 1 \ \text{ such that for all } \ c \ \text{ in } \ [1-\delta, 1+\delta], \\ \text{we can find numbers } \ c_1, c_2, c_3 \ \text{ such that} \\ \phi(cx) \le c_1\phi(x) + c_2|x| + c_3, \quad \text{for all real } \ x, \end{array} \tag{2.4}$$

then the assumption (2.3) is satisfied whenever $D_\phi(\theta,\alpha) < \infty$, where $D_\phi(\theta,\alpha)$ stands for the ϕ-divergence between $\mathbb{P}_\theta$ and $\mathbb{P}_\alpha$, refer to [14, Lemma 3.2]. Also the real convex functions $\phi(\cdot)$ (2.2), associated with the class of power divergences, all satisfy the condition (2.3), including all standard divergences.
According to [33], under the strict convexity and the differentiability of the function ϕ, it holds

$$\phi(t) \ge \phi(s) + \phi'(s)(t-s), \tag{2.5}$$

where the equality holds only for $s = t$. Let θ, θ_0, η and η_0 be fixed and put $t = \mathbb{P}_{\theta,\eta}(x)/\mathbb{P}_{\theta_0,\eta_0}(x)$ and $s = \mathbb{P}_{\theta,\eta}(x)/\mathbb{P}_{\alpha,\eta}(x)$ in (2.5) and then integrate with respect to $\mathbb{P}_{\theta_0,\eta_0}$. Under (2.3), this gives

$$\begin{aligned} D_\phi(\theta,\theta_0) &= \int \phi\left(\frac{\mathbb{P}_{\theta,\eta}(x)}{\mathbb{P}_{\theta_0,\eta_0}(x)} \right) \, \mathrm{d}\mathbb{P}_{\theta_0,\eta_0}(x) \\ &= \sup_{\alpha \in \Theta} \int m(\theta,\alpha,\eta) \, \mathrm{d}\mathbb{P}_{\theta_0,\eta_0}, \end{aligned} \tag{2.6}$$

where $m(\theta,\alpha,\eta) : x \mapsto m(\theta,\alpha,\eta,x)$ and

$$\begin{aligned} m(\theta,\alpha,\eta,x) &:= \int \phi'\left(\frac{\mathbb{P}_{\theta,\eta}}{\mathbb{P}_{\alpha,\eta}} \right) \, \mathrm{d}\mathbb{P}_{\theta,\eta} \\ &\quad - \left[\frac{\mathbb{P}_{\theta,\eta}(x)}{\mathbb{P}_{\alpha,\eta}(x)} \phi'\left(\frac{\mathbb{P}_{\theta,\eta}(x)}{\mathbb{P}_{\alpha,\eta}(x)} \right) - \phi\left(\frac{\mathbb{P}_{\theta,\eta}(x)}{\mathbb{P}_{\alpha,\eta}(x)} \right) \right]. \end{aligned} \tag{2.7}$$

The supremum in (2.6) is unique and is attained in $\alpha = \theta_0$, independently upon the value of θ. In the sequel we use the notation

$$\mathbb{P}_n\psi(\gamma) := \frac{1}{n}\sum_{i=1}^n \psi(X_i),$$

where $\psi(\cdot)$ is a deterministic measurable function and $\mathbb{P}_n$ is the empirical measure associated with these random variables is defined as placing mass $1/n$ on each of the observations X_i, $i = 1, \ldots, n$, i.e.,

$$\mathbb{P}_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

where δ_x denotes Dirac measure at point $x \in \mathcal{X}$. An estimator of (θ_0, η_0) has the form

$$(\widehat{\alpha}_\phi(\theta), \widehat{\eta}) = \arg\max \mathbb{P}_n m(\theta, \alpha, \eta), \tag{2.8}$$

where $m(\theta, \alpha, \eta)$ is the function defined in (2.7).
Formula (2.8) defines a family of M-estimators indexed by the function ϕ specifying the divergence and by some instrumental value of the parameter θ, called here escort parameter. The term "M-estimation" refers to a general method of estimation, where the estimators are obtained by maximizing (or minimizing) certain criterion functions. The most widely used M-estimators include maximum-likelihood (MLE), ordinary least-squares (OLS), and least absolute deviation estimators. The choices of ϕ and θ represent a major feature of the estimation procedure, since they induce efficiency and robustness properties. Asymptotic properties of the above estimators can be handled through the general theory of M-estimators for semiparametric models, see for instance [31].

2.2. Estimation procedures

Estimation in general semiparametric models by Hellinger distance was studied in [56]. Naturally, extension of theses results, is the class of estimators of θ_0, called "profile dual ϕ-divergence estimators" (pDϕDE's), is defined by

$$\widehat{\alpha}_\phi(\theta) := \arg\sup_{\alpha\in\Theta} \mathbb{P}_n m(\theta, \alpha, \widehat{\eta}), \quad \theta \in \Theta, \tag{2.9}$$

The class of estimators $\widehat{\alpha}_\phi(\theta)$ satisfies

$$\mathbb{P}_n \frac{\partial}{\partial\alpha} m(\theta, \widehat{\alpha}_\phi(\theta), \widehat{\eta}) = 0. \tag{2.10}$$

The theory developed in this paper is general enough to deal with the case that $(\widehat{\alpha}_\phi(\theta), \widehat{\eta})$ is not the exact maximizer. Instead of (2.10), we can only assume the following "nearly-maximizing" condition

$$\mathbb{P}_n \frac{\partial}{\partial\alpha} m(\theta, \widehat{\alpha}_\phi(\theta), \widehat{\eta}) = o_{\mathbb{P}^*}(n^{-1/2}). \tag{2.11}$$

We will use the following notation

$$\Psi(\theta, \alpha, \eta) = \mathbb{P}_{\theta_0,\eta_0} \frac{\partial}{\partial\alpha} m(\theta, \alpha, \eta) \tag{2.12}$$

where $\mathbb{P}_{\theta_0,\eta_0}\psi$ is the customary operator notation defined as $\int_{\mathcal{X}} \psi(x) d\mathbb{P}_{\theta_0,\eta_0}(x)$, and

$$\Psi_n(\theta, \alpha, \eta) = \mathbb{P}_n \frac{\partial}{\partial\alpha} m(\theta, \alpha, \eta). \tag{2.13}$$

The true value θ_0 of θ then satisfies $\Psi(\theta, \theta_0, \eta_0) = 0$, and

$$\widehat{\alpha}_\phi(\theta) = \arg\min_{\alpha\in\Theta} \left\| \Psi_n(\theta, \alpha, \widehat{\eta}_n) \right\|,$$

where $\|\cdot\|$ denotes the Euclidean norm.

3. Asymptotics

3.1. Consistency

The consistency of a semiparametric M-estimator $\widehat{\alpha}_\phi(\theta)$ can be obtained using general results available in the literature. The consistency of general M-estimators have been investigated at length by a number of authors, among whom we may cite [41, Theorem 2.1.], [51, Corollary 3.2.3.] and [19, Theorem 1.]. In particular, we will use the results of [19] who proposed conditions under which a parameter estimator that is defined via an estimating equation depending on some nonparametric nuisance functions, is consistent and asymptotically normal, which plays an instrumental role in proving consistency of our estimators. We now state some general conditions that will be used throughout the whole paper.

(C.1) θ_0 is the unique solution to $\Psi(\theta,\alpha,\eta_0(\cdot;\alpha))=0$ in the parameter space Θ;

(C.2) $\hat{\eta}_n$ is an estimator of η_0 such that

$$\|\hat{\eta}_n-\eta_0\|=o_{\mathbb{P}^*}(1);$$

(C.3) for every sequence $\{\delta_n\}\downarrow 0$

$$\sup_{\alpha\in\Theta,\|\eta-\eta_0\|\leq\delta_n}\frac{|\Psi_n(\theta,\alpha,\eta(\cdot;\theta))-\Psi(\theta,\alpha,\eta_0(\cdot;\alpha))|}{1+|\Psi_n(\theta,\alpha,\eta(\cdot;\theta))|+|\Psi(\theta,\alpha,\eta_0(\cdot;\alpha))|}=o_{\mathbb{P}^*}(1). \tag{3.1}$$

Theorem 1. Assume that conditions (C.1-3) hold. Then, $\widehat{\alpha}_\phi(\theta)$ satisfying

$$\Psi_n(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))=o_{\mathbb{P}^*}(1),$$

converges in outer probability to θ_0.

We need the following definitions, refer to [50] and [51] among others. If $\mathcal{F}$ is a class of functions for which, we have almost surely,

$$\|\mathbb{P}_n-\mathbb{P}\|_{\mathcal{F}}=\sup_{f\in\mathcal{F}}|\mathbb{P}_nf-\mathbb{P}f|\to 0,$$

then we say that $\mathcal{F}$ is a $\mathbb{P}$-Glivenko-Cantelli class of functions. If $\mathcal{F}$ is a class of functions for which

$$\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-\mathbb{P})\to\mathbb{G}\ \text{ in }\ \ell^\infty(\mathcal{F}),$$

where $\mathbb{G}$ is a mean-zero $\mathbb{P}$-Brownian bridge process with (uniformly-) continuous sample paths with respect to the semi-metric $\rho_{\mathbb{P}}(f,g)$, defined by

$$\rho^2_{\mathbb{P}}(f,g)=Var_{\mathbb{P}}(f(X)-g(X)),$$

then we say that $\mathcal{F}$ is a $\mathbb{P}$-Donsker class of functions. Here

$$\ell^\infty(\mathcal{F})=\left\{v:\mathcal{F}\mapsto\mathbb{R}\middle|\|v\|_{\mathcal{F}}=\sup_{f\in\mathcal{F}}|v(f)|<\infty\right\}$$

and $\mathbb{G}$ is a $\mathbb{P}$-Brownian bridge process on $\mathcal{F}$ if it is a mean-zero Gaussian process with covariance function

$$\mathbb{E}(\mathbb{G}(f)\mathbb{G}(g))=\mathbb{P}fg-(\mathbb{P}f)(\mathbb{P}g).$$

Condition (C.3) of Theorem 1 is implied by the following condition

(C.3)' $$\sup_{\theta\in\Theta,\|\eta-\eta_0\|\leq\delta_n}|\Psi_n(\theta,\alpha,\eta(\cdot;\theta))-\Psi(\theta,\eta_0(\cdot;\alpha))|=o_{\mathbb{P}^*}(1).$$

Observe that

$$\Psi_n(\theta,\alpha,\eta)-\Psi(\theta,\alpha,\eta)=(\mathbb{P}_n-\mathbb{P}_{\theta_0,\eta_0})\,\varphi\,(\theta,\alpha,\eta),$$

with

$$\varphi\,(\theta,\alpha,\eta)=\frac{\dot{p}_{\alpha,\eta}}{\mathbb{P}_{\alpha,\eta}}\left(\frac{\mathbb{P}_{\theta,\eta}}{\mathbb{P}_{\alpha,\eta}}\right)^2\phi''\left(\frac{\mathbb{P}_{\theta,\eta}}{\mathbb{P}_{\alpha,\eta}}\right). \tag{3.2}$$

Let θ be fixed,

$$\mathcal{F}_\theta = \{\varphi(\theta, \alpha, \eta) : \alpha \in \Theta, \eta \in \mathcal{H}\}$$

denotes the class of measurable functions indexed by (α, η). By modern empirical process theory presented in [51] for example, condition (C.3)' will be satisfied when $\mathcal{F}_\theta$ is P-Glivenko-Cantelli. In the case of power divergence the class $\mathcal{F}_\theta$ will reduce to

$$\mathcal{F}_\theta = \left\{ \frac{\dot{p}_{\alpha,\eta}}{\mathbb{P}_{\alpha,\eta}} \left(\frac{\mathbb{P}_{\theta,\eta}}{\mathbb{P}_{\alpha,\eta}} \right)^\gamma : \alpha \in \Theta, \eta \in \mathcal{H} \right\}.$$

By application of the Glivenko-Cantelli preservation properties, we can show that $\mathcal{F}_\theta$ is P-Glivenko-Cantelli, by showing first, that the classes $\{\dot{p}_{\alpha,\eta}\}$, $\{\mathbb{P}_{\alpha,\eta}\}$ and $\{\mathbb{P}_{\theta,\eta}\}$ are P-Glivenko-Cantelli, for more details on the subject refer to [49].

3.2. Asymptotic normality

Let

$$\mathcal{H}_0 = \{\eta(x;\theta) : x \in \mathcal{X}, \theta \in \Theta_0\}$$

be a collection of functions that are continuously differentiable in θ for all $x \in \mathcal{X}$ with bounded derivative matrices $\{\dot{\eta}(\cdot;\theta)\}$, where $\Theta_0 \subset \Theta$ is a neighborhood of θ_0. Suppose that $\widehat{\alpha}_\phi(\theta)$ satisfying

$$\Psi_n(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta))) = o_{\mathbb{P}^*}(n^{-1/2}),$$

is a consistent estimator of θ_0 that is the unique solution to the equation

$$\Psi(\theta, \alpha, \eta_0(\cdot;\alpha)) = 0$$

in Θ, and that $\hat{\eta}_n \in \mathcal{H}_0$ is an estimator of $\eta_0 \in \mathcal{H}_0$ satisfying

$$\|\hat{\eta}_n - \eta_0\| = o_{\mathbb{P}^*}(n^{-\beta}),$$

for some $\beta > 0$. Suppose the following four conditions are satisfied:

(A.1) (Stochastic equicontinuity.)

$$\frac{|n^{1/2}(\Psi_n - \Psi)(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta))) - n^{1/2}(\Psi_n - \Psi)(\theta, \theta_0, \eta_0(\cdot;\theta_0))|}{1 + n^{1/2}|\Psi_n(\widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta)))| + n^{1/2}|\Psi(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta)))|} = o_{\mathbb{P}^*}(1)\ .$$

(A.2) $n^{1/2}\Psi_n(\theta, \theta_0, \eta_0(\cdot;\theta_0)) = o_{\mathbb{P}^*}(1)$.

(A.3) (Smoothness.) (a) If $\beta = 1/2$, the function $\Psi(\theta, \alpha, \eta(\cdot;\theta)) : \Theta_0 \times \mathcal{H}_0 \to \mathbb{R}^d$ is Fréchet differentiable at $(\theta_0, \eta_0(\cdot;\theta_0))$, i.e., there exists a continuous $d \times d$ matrix $\dot{\Psi}_1(\theta_0, \eta_0(\cdot;\theta_0))$ and a continuous linear functional $\dot{\Psi}_2(\theta, \theta_0, \eta_0(\cdot;\theta_0))$ such that

$$\begin{aligned} &|\Psi(\theta, \alpha, \eta(\cdot;\theta)) - \Psi(\theta, \theta_0, \eta_0(\cdot;\theta_0)) \\ &\qquad - \{\dot{\Psi}_1(\theta, \theta_0, \eta_0(\cdot;\theta_0)) + \dot{\Psi}_2(\theta, \theta_0, \eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]\}(\alpha - \theta_0) \\ &\qquad - \dot{\Psi}_2(\theta, \theta_0, \eta_0(\cdot;\theta_0))[(\eta - \eta_0)(\cdot;\theta_0)]| \\ &= o(|\alpha - \theta_0|) + o(\|\eta - \eta_0\|)\,; \end{aligned} \tag{3.3}$$

or (b) if $0 < \beta < 1/2$, for some $\xi > 1$ satisfying $\xi\beta > 1/2$ we have

$$\begin{aligned} &|\Psi(\theta, \alpha, \eta(\cdot;\theta)) - \Psi(\theta, \theta_0, \eta_0(\cdot;\theta_0)) \\ &\qquad - \{\dot{\Psi}_1(\theta, \theta_0, \eta_0(\cdot;\theta_0)) + \dot{\Psi}_2(\theta, \theta_0, \eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]\}(\alpha - \theta_0) \\ &\qquad - \dot{\Psi}_2(\theta, \theta_0, \eta_0(\cdot;\theta_0))[(\eta - \eta_0)(\cdot;\theta_0)]| \\ &= o(|\alpha - \theta_0|) + O(\|\eta - \eta_0\|^\xi)\,. \end{aligned} \tag{3.4}$$

Here the subscripts 1 and 2 correspond to the first and the second arguments in $\Psi(\cdot,\cdot)$, respectively, and we assume that the matrix

$$A = -\dot{\Psi}_1(\theta, \theta_0, \eta_0(\cdot;\theta_0)) - \dot{\Psi}_2(\theta, \theta_0, \eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]$$

is nonsingular.

(A.4)
$$n^{1/2}\dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[(\hat{\eta}_n-\eta_0)(\cdot;\theta_0)] = o_{\mathbb{P}^*}(1).$$

The main result of the present paper is given in the following theorem.

Theorem 2. Suppose that conditions (A.1-4) hold. Then $\widehat{\alpha}_\phi(\theta)$ is $n^{1/2}$-consistent and further we have

$$\begin{aligned} n^{1/2}(\widehat{\alpha}_\phi(\theta)-\theta_0) &= A^{-1}n^{1/2}\Big\{(\Psi_n-\Psi)(\theta,\theta_0,\eta_0(\cdot;\theta_0)) \\ &\quad + \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[(\hat{\eta}_n-\eta_0)(\cdot;\theta_0)]\Big\} + o_{\mathbb{P}^*}(1). \end{aligned} \tag{3.5}$$

4. Examples

In this section, we consider three examples of the semiparametric models (2.1): the symmetric location model, generalized logistic models and a scale mixture model. In each case, we will demonstrate the construction of the pDϕD estimator defined by (2.10) for the parameters of interest. In all the following we will consider the class of "power divergences" defined in (2.2).

4.1. Symmetric location

Assume that the data $X_1,\ldots,X_n \in \mathbb{R}$ are i.i.d. and satisfy the model

$$X = \theta + \varepsilon,$$

where the center θ is the parameter to be estimated, and the error ε has a symmetric (about zero) continuous bounded density η. Then the semiparametric model under our consideration here is

$$\mathcal{P} = \{\mathbb{P}_{\theta,\eta}(x) = \eta(x-\theta):\ \theta\in\mathbb{R},\ \eta\in\mathcal{H}\}$$

where

$$\begin{aligned} \mathcal{H} &= \Big\{\eta:\ \eta>0,\ \int\eta(x)\,\mathrm{d}x = 1,\ \eta(-x)=\eta(x), \\ &\qquad \eta \text{ is absolutely continuous a.e. with } \int\frac{(\eta'(x))^2}{\eta(x)}\,\mathrm{d}x<\infty\Big\}. \end{aligned}$$

Since $\eta(x)=\mathbb{P}_{\theta,\eta}(x+\theta)$, intuitively we can construct an estimator of η as

$$\hat{p}(x+\tilde{\theta}) := \frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{x+\tilde{\theta}-X_i}{h_n}\right),$$

where $\hat{p}$ is the kernel density estimator of $\mathbb{P}_{\theta,\eta}$ based on $X_1,\ldots,X_n$ and $\tilde{\theta}$ is a preliminary estimator of θ. See [55]. Suppose that the model consists of all densities $x\mapsto\eta(x-\theta)$ with $\theta\in\mathbb{R}$ and the density $\eta(\cdot)$ symmetric about 0 with finite Fisher information.

$$\mathcal{F}_\theta = \left\{-\frac{\eta'(x-\alpha)}{\eta(x-\alpha)}\left(\frac{\eta(x-\theta)}{\eta(x-\alpha)}\right)^\gamma : \alpha\in\Theta\right\}. \tag{4.1}$$

4.2. Regression

Consider the regression model $Y=m(X)+\varepsilon$ where the regression function is parameterized such that

$$Y = g(X,\theta_0)+\varepsilon, \tag{4.2}$$

where X is a $q\times 1$ vector of covariates, g is the regression function of known form, θ_0 is a vector of unknown parameters, and ε is the error independent of X with mean zero and finite variance. The model is semiparametric in the sense that the regression function is parametric and the error term distribution is nonparametric. Let X and ϵ be independent random vectors with densities $g(\cdot)$ and $f(\cdot)$ respectively and suppose that

$$Y = r(\theta, X) + \epsilon, \tag{4.3}$$

for function $r(\theta, X)$ that is known up to θ. Semiparametric versions are obtained by letting the distribution of ϵ range over all distributions on the real line with mean zero, or, alternatively, over all distributions that are symmetric about zero. Thus, the observation (X, Y) has a density

$$\eta_0(x, y - r(\theta, x)) = g(x)f(y - r(\theta, x)).$$

$$\mathcal{F}_\theta = \left\{ -\frac{\partial}{\partial \alpha} r(\alpha, x) \frac{f'(y - r(\alpha, x))}{f(y - r(\alpha, x))} \left(\frac{f(y - r(\theta, x))}{f(y - r(\alpha, x))} \right)^\gamma : \alpha \in \Theta \right\}. \tag{4.4}$$

4.3. Generalized logistic models

Following [55], suppose Y is a binary response variable and X is the associated covariate; then the (prospective) logistic regression model is of the form

$$\mathbb{P}(Y = 1|X = x) = \frac{\exp[a' + bx]}{1 + \exp[a' + bx]},$$

where a' and b are parameters and the marginal distribution of X is not specified. In case-control studies, data are collected retrospectively in the sense that for samples of subjects having $Y = 1$ ('case') and having $Y = 0$ ('control'), the value x of X is observed. More specifically, suppose $X_1, \ldots, X_n$ is a random sample from $F(x|Y = 1)$ and, independently of the X_i's, suppose $Z_1, \ldots, Z_m$ is a random sample from $F(x|Y = 0)$. If $\pi = \mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0)$ and $f(x|Y = i)$ is the conditional density of X given $Y = i$, $i = 0, 1$, then it follows from Bayes rule that

$$f(x|Y = 1) = f(x|Y = 0) \exp[a + bx],$$

where

$$a = a' + \log[(1 - \pi)/\pi].$$

In other words, we observe two independent samples

$$\begin{aligned} &Z_1, \ldots, Z_m \text{ i.i.d.} \sim \eta(x), \\ &X_1, \ldots, X_n \text{ i.i.d.} \sim f(x) = f_{\theta,\eta}(x) = \eta(x) \exp[a + r(x)b]. \end{aligned} \tag{4.5}$$

We are concerned with estimation of parameter $\theta = (a, b)$ when η is unknown (the nuisance parameter).

5. Exchangeably weighted bootstraps of minimum profile divergence

Bootstrap samples were introduced and first investigated in [25]. Since this seminal paper, bootstrap methods have been proposed, discussed, investigated and applied in a huge number of papers in the literature. Being one of the most important ideas in the practice of statistics, the bootstrap also introduced a wealth of innovative probability problems, which in turn formed the basis for the creation of new mathematical theories. The asymptotic theory of the bootstrap with statistical applications has been reviewed in the books among others [26], [28] and [31]. 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 come. 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. Roughly speaking, it is known that the bootstrap works in the i.i.d. case if and only if the central limit theorem holds for the random variable under consideration. For further discussion we refer the reader to the landmark paper by [27]. 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 $f(\cdot)$,

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

where W_{ni}'s are the bootstrap weights defined on the probability space $(\mathcal{W}, \Omega, \mathbb{P}_W)$. Following [21], assume there exists an

$$H(\theta,\eta) = (h_1(\theta,\eta),\ldots,h_p(\theta,\eta))^\top,$$

where each $h_j(\theta,\eta) \in \mathcal{H}$, such that for any $h \in \mathcal{H}$

$$\mathbb{E}_{\theta,\eta}[m_{12}(\theta,\alpha,\eta)[h] - m_{22}(\theta,\alpha,\eta)[H,h]] = 0,$$

where

$$m_{12}(\theta,\alpha,\eta)[h] = \frac{\partial}{\partial t} m_1(\theta,\alpha,\eta(t))[h]\bigg|_{t=0},$$

and

$$m_{22}(\theta,\alpha,\eta)[h] = \frac{\partial}{\partial t} m_2(\theta,\alpha,\eta(t))[h]\bigg|_{t=0}.$$

We define the function

$$\widetilde{m}(\theta,\alpha,\eta) = m_1(\theta,\alpha,\eta) - m_2(\theta,\alpha,\eta)[H(\theta,\eta)].$$

In view of (2.9), the bootstrap estimator can be rewritten as

$$(\widehat{\alpha}_\phi^*(\theta), \widehat{\eta}_n^*) := \arg\sup_{\alpha\in\boldsymbol{\Theta},\eta\in\mathcal{H}} \mathbb{P}_n^* m(\theta,\alpha,\eta). \tag{5.1}$$

The definition of $\widehat{\alpha}_\phi^*(\theta)$, defined in (5.1), implies that

$$\mathbb{P}_n^* \frac{\partial}{\partial\alpha} m(\theta, \widehat{\alpha}_\phi^*(\theta), \widehat{\eta}_n^*) = 0. \tag{5.2}$$

One can see also that

$$\mathbb{P}_n^* \widetilde{m}(\theta, \widehat{\alpha}_\phi^*(\theta), \widehat{\eta}_n^*) = 0. \tag{5.3}$$

The bootstrap weights W_{ni}'s are assumed to belong to the class of exchangeable bootstrap weights introduced in [43]. In the sequel, the transpose of a vector $\mathbf{x}$ will be denoted by $\mathbf{x}^\top$. We shall assume the following conditions.

W.1 *The vector $W_n = (W_{n1},\ldots,W_{nn})^\top$ is exchangeable for all $n = 1,2,\ldots$, i.e., for any permutation $\pi = (\pi_1,\ldots,\pi_n)$ of $(1,\ldots,n)$, the joint distribution of*

$$\pi(W_n) = (W_{n\pi_1},\ldots,W_{n\pi_n})^\top$$

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

$$\lim\sup_{n\to\infty} \|W_{n1}\|_{2,1} \leq C < \infty,$$

where

$$\|W_{n1}\|_{2,1} = \int_0^\infty \sqrt{\mathbb{P}_W(W_{n1} \geq u)}\,du.$$

W.4

$$\lim_{\lambda\to\infty} \lim \sup_{n\to\infty} \sup_{t\geq\lambda} t^2 \mathbb{P}_W(W_{n1} > t) = 0.$$

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 [43, 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 [43], [51] and [31].

There exist two sources of randomness for the bootstrapped quantity, i.e., $\widehat{\alpha}^*_\phi(\theta)$: 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. Following [21] and [53], we shall view X_i as the i-th coordinate projection from the canonical probability space $(\mathcal{X}^\infty, \mathcal{A}^\infty, \mathbb{P}^\infty_{\theta_0,\eta_0})$ onto the i-th copy of $\mathcal{X}$. For the joint randomness involved, the product probability space is defined as

$$(\mathcal{X}^\infty, \mathcal{A}^\infty, \mathbb{P}^\infty_{\theta_0,\eta_0}) \times (\mathcal{W}, \Omega, \mathbb{P}_W) = (\mathcal{X}^\infty \times \mathcal{W}, \mathcal{A}^\infty \times \Omega, \mathbb{P}^\infty_{\theta_0,\eta_0} \times \mathbb{P}_W).$$

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

$$\mathbb{P}_{XW} = \mathbb{P}_{\theta_0,\eta_0} \times \mathbb{P}_W.$$

Given a real-valued function Δ_n defined on the above product probability space, e.g. $\widehat{\alpha}^*_\phi(\theta)$, we say that Δ_n is of an order $o_{\mathbb{P}_W}(1)$ in $\mathbb{P}_{\theta_0,\eta_0}$-probability if, for any $\epsilon, \eta > 0$, as $n \to 0$,

$$\mathbb{P}_{\theta_0,\eta_0}\{\mathbb{P}_{W|X}(|\Delta_n| > \epsilon) > \eta\} \quad \longrightarrow \quad 0, \tag{5.4}$$

and that Δ_n is of an order $O_{\mathbb{P}_W}(1)$ in $\mathbb{P}_{\theta_0,\eta_0}$-probability if, for any $\eta > 0$, there exists a $0 < M < \infty$ such that, as $n \to 0$,

$$\mathbb{P}_{\theta_0,\eta_0}\{\mathbb{P}_{W|X}(|\Delta_n| \geq M) > \eta\} \longrightarrow 0. \tag{5.5}$$

We shall say a class of functions $\mathcal{H} \in M(\mathbb{P}_{\theta_0,\eta_0})$ 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

$$(\delta_{X_i} - \mathbb{P}_{\theta_0,\eta_0})$$

by

$$(W_{ni} - 1)\delta_{X_i}.$$

It is known that $\mathcal{H} \in M(\mathbb{P}_{\theta_0,\eta_0})$, e.g., if $\mathcal{H}$ is countable, or if $\{\mathbb{P}_n\}^\infty_n$ are stochastically separable in $\mathcal{H}$, or if $\mathcal{H}$ is image admissible Suslin; see [27, pages 853 and 854].

For any fixed $\delta_n > 0$, define a class of function $\mathcal{S}_n$ as

$$\mathcal{S}_n := \mathcal{S}_n(\delta_n) = \left\{ \frac{\widetilde{m}(\theta,\theta_0,\eta) - \widetilde{m}(\theta,\theta_0,\eta_0)}{\|\eta-\eta_0\|_{\mathcal{H}}} : \|\eta-\eta_0\|_{\mathcal{H}} \leq \delta_n \right\}.$$

We define a

$$\mathcal{C}_n(\delta_n) := \mathbf{\Theta}_\delta \times \mathcal{H}_{\delta_n}.$$

To state our result concerning the asymptotic normality, we shall assume the following additional conditions.

(A.8) The tail probability condition:

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \sup_{t\geq\lambda} t^2 \mathbb{P}(S_n(X_1) > t) = 0,$$

where $S_n(x)$ is the envelop function of the class $\mathcal{S}_n$, that is,

$$S_n(x) = \sup_{\|\eta-\eta_0\|_{\mathcal{H}}\leq\delta_n} \left| \frac{\widetilde{m}(\theta_0,\eta) - \widetilde{m}(\theta_0,\eta_0)}{\|\eta-\eta_0\|_{\mathcal{H}}} \right|;$$

(A.9) The class $\dot{\mathcal{I}}_n \in M(\mathbb{P}_{\theta_0,\eta_0}) \cap L_2(\mathbb{P}_{\theta_0,\eta_0})$ is $\mathbb{P}$-Donsker, where

$$\dot{\mathcal{I}}_n := \{\partial\widetilde{m}(\theta,\alpha,\eta)/\partial\alpha : (\alpha,\eta) \in \mathcal{C}_n\};$$

(A.10)

$$\|\widehat{\eta}_n^* - \eta_0\|_{\mathcal{H}} = o_{\mathbb{P}_W}(n^{-1/4}),$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probabbility.

Conditions (A.4) and (A.5) ensure that the "size" of the function class $\dot{\mathcal{I}}_n$ is reasonable so that the bootstrapped empirical processes $\mathbb{G}_n^* \equiv \sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ indexed by $\dot{\mathcal{I}}_n$, has a limiting process conditional on the original observations, we refer for instance to [43, Theorem 2.2]. The main result to be proved here may now be stated precisely as follows.

Theorem 3. Assume that $\widehat{\alpha}_\phi(\theta)$ and $\widehat{\alpha}_\phi^*(\theta)$ fullfil (2.10) and (5.2), respectively. In addition suppose that

$$\widehat{\alpha}_\phi(\theta) \xrightarrow{\mathbb{P}_{\theta_0,\eta_0}} \theta_0 \text{ and } \widehat{\alpha}_\phi^*(\theta) \xrightarrow{\mathbb{P}_W} \theta_0 \text{ in } \mathbb{P}_{\theta_0,\eta_0}\text{-probability.}$$

Assume that conditions (A.1–9) and W.1–W.5 hold. Then we have

$$\|\widehat{\alpha}_\phi^*(\theta) - \theta_0\| = O_{\mathbb{P}_W}(n^{-1/2}) \tag{5.6}$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability. Furthermore,

$$\sqrt{n}(\widehat{\alpha}_\phi^*(\theta) - \widehat{\alpha}_\phi(\theta)) = -\Gamma_1^{-1}\mathbb{G}_n^*\widetilde{m}(\theta,\theta_0,\eta_0) + o_{\mathbb{P}_W}(1) \tag{5.7}$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability. Consequently,

$$\sup_{x\in\mathbb{R}^d} \left| \mathbb{P}_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\alpha}_\phi^*(\theta) - \widehat{\alpha}_\phi(\theta)) \leq x) - \mathbb{P}(N(0,\Sigma) \leq x) \right| = o_{\mathbb{P}_{\theta_0,\eta_0}}(1), \tag{5.8}$$

where "$\leq$" is taken componentwise and "c" is given in W.5, whose value depends on the used sampling scheme, and

$$\Sigma \equiv \Gamma_1^{-1} V_1 \Gamma_1^{-1}.$$

Thus, we have

$$\sup_{x\in\mathbb{R}^d} \left| \mathbb{P}_{W|\mathcal{X}_n}((\sqrt{n}/c)(\widehat{\alpha}_\phi^*(\theta) - \widehat{\alpha}_\phi(\theta)) \leq x) - \mathbb{P}_{\theta_0,\eta_0}(\sqrt{n}(\widehat{\alpha}_\phi(\theta) - \theta_0) \leq x) \right| \xrightarrow{\mathbb{P}_{\theta_0,\eta_0}} 0. \tag{5.9}$$

For application of this result, one can refer to [13], [12], [8, 9] and [1, 2, 3].

Remark 5.1. Notice that the choice of weights depends on the problem at hand : accuracy of 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, we may refer to [35] and the references therein for more details. [4] indicate that the area where the weighted bootstrap clearly performs better than the classical bootstrap is in term of coverage accuracy.

Remark 5.2. 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 [43, Examples 3.1 and 3.4].

Remark 5.3. In order to extract methodological recommendations for the use of an appropriate divergence, it will be interesting to conduct an extensive Monte Carlo experiments for several divergences or investigate theoretically the problem of the choice of the divergence which leads to an "*optimal*" (in some sense) estimate in terms of efficiency and robustness, which would go well beyond the scope of the present paper. An other challenging task is how to choose the bootstrap weights for a given divergence in order to obtain, for example, an efficient estimator.

Bootstrap weights

Let us present some examples of the bootstrap weights satisfying the conditions W.1-W.5, we can refer to [43] and [20] for further details. More precisely, the following examples are provided in this compressed form in [20], we have included some minor changes necessary for our setting.

Example 5.1 (i.i.d.-Weighted Bootstraps)**.** In this example, the bootstrap weights are defined as

$$W_{ni} = \omega_i/\overline{\omega}_n,$$

where $\omega_1, \omega_2, \ldots, \omega_n$ are i.i.d. positive r.v.s. with $\|\omega_1\|_{2,1} < \infty$, where

$$\|W_{n1}\|_{2,1} = \int_0^\infty \sqrt{\mathbb{P}_W(W_{n1} \geq u)}du,$$

$$\overline{\omega}_n = \sum_{i=1}^n \omega_i.$$

Thus, we can choose $\omega_i \sim \text{Exponential}(1)$ or $\omega_i \sim \text{Gamma}(4,1)$. The former corresponds to the Bayesian bootstrap. The multiplier bootstrap is often thought to be a smooth alternative to the nonparametric bootstrap; see [34]. The value of c^2 is calculated as

$$Var(\omega_1)/(E\omega_1)^2.$$

Example 5.2 (Efron's bootstrap)**.** As already mentioned, the weights for the Efron bootstrap satisfy the conditions W.1-W.5 with $c^2 = 1$ and are

$$W_n \sim \text{Multinomial}(n; n^{-1}, \ldots, n^{-1}).$$

Example 5.3 (The delete-h Jackknife)**.** In the delete-h jackknife, see [57], the bootstrap weights are generated by permuting the deterministic weights

$$w_n = \left\{\frac{n}{n-h}, \ldots, \frac{n}{n-h}, 0, \ldots, 0\right\} \quad \text{with} \sum_{i=1}^n w_{ni} = n.$$

Specifically, we have $W_{nj} = w_{nR_n(j)}$ where $R_n(\cdot)$ is a random permutation uniformly distributed over $\{1, \ldots, n\}$. In Condition W.5, $c^2 = h/(n-h)$. Thus, we need to choose $h/n \to \alpha \in (0,1)$ such that $\varrho > 0$. Therefore, the usual jackknife with $h = 1$ is inconsistent for estimating the distribution.

Let us recall some examples from [36].

Example 5.4. The $m(n)$ out of n-bootstrap weights

$$W_{ni} = m(n)^{1/2}\left(\frac{1}{m(n)}M_{ni} - \frac{1}{n}\right)$$

are given by a multinomial distributed random variable $(M_{n1}, \dots, M_{n,n})$ with sample size

$$m(n) = \sum_{i=1}^{n} M_{ni}$$

and equal success probability. In this case, the conditions W.1-W.5 are valid, (details of the proof are given in [37, (8.37)-(8.46)]).

Example 5.5. The $m(n)$-double bootstrap can be described by the weights

$$W_{ni} = \frac{m(n)^{1/2}}{\sqrt{2}}\left(\frac{1}{m(n)}M'_{ni} - \frac{1}{n}\right)$$

Here $(M'_{n1}, \dots, M'_{nn})$ denotes a conditional multinomial distributed variable with sample size

$$m(n) = \sum_{i=1}^{n} M_{ni}$$

and success probability $M_{ni}/m(n)$ for the i-th cell given by the first example, (details of this example are discussed in Lemma 6.2 of [36]).

Remark 5.4. As was pointed out in [43], the aforementioned bootstraps are "smoother" in some sense than the multinomial bootstrap since they put some (random) weight at all elements in the sample, whereas the multinomial bootstrap puts positive weight at about

$$1 - (1 - n^{-1})^n \to 1 - e^{-1} = 0.6322$$

proportion of each element of the sample, on the average. Notice that when $\omega_i \sim$ Gamma$(4, 1)$ so that the W_{ni}/n are equivalent to four-spacings from a sample of $4n - 1$ Uniform(0,1) random variables. In [54] and [52], it was noticed that, in addition to being four times more expensive to implement, the choice of four-spacings depends on the functional of interest and is not universal.

6. Class of Functional M-Estimator processes

Let $\ell^\infty(\mathcal{H})$ denote the set of bounded functions from $\mathcal{H}$ to the real line $\mathbb{R}$, for some set $\mathcal{H}$, and let $\|\cdot\|_{\mathcal{H}}$ denote the uniform norm on $\ell^\infty(\mathcal{H})$. Let $\psi(\gamma; \cdot)$ be a γ-indexed operator from $\mathcal{H}$ to some subset $\mathcal{F}(\gamma)$ of $L_2(\mathbb{P}_\gamma)$ for each $\gamma \in \mathbf{\Theta}$. Define the set

$$\mathcal{F}(\mathbf{\Theta}) = \bigcup_{\gamma\in\mathbf{\Theta}} \mathcal{F}(\gamma).$$

For simplicity of notation, we omit $\mathbf{\Theta}$ in $\mathcal{F}(\mathbf{\Theta})$ and simply write $\mathcal{F}$. A functional Z-estimator for θ_0 is a sequence of estimates $\widehat{\theta}_n$ which makes the "scores" $\mathbb{P}_n\psi(\gamma; \cdot)(h), h \in \mathcal{H}$, approximately zero, that is,

$$\|\mathbb{P}_n\psi(\widehat{\theta}_n; \cdot)\|_{\mathcal{H}} = o_{\mathbb{P}^*}(n^{-1/2}), \tag{6.1}$$

where $\mathbb{P}^*$ denotes the outer probability of $\mathbb{P}^\infty$ (see, e.g., [51] and [31] for more details on outer probability measures). According to equation (6.1) the estimator $\widehat{\theta}_n$ depends on the underlying function $\psi(\cdot)$ and therefore it can be denoted by $\widehat{\theta}_{n;\psi}$. Varying the function $\psi(\cdot)$ over a class of measurable functions Ψ leads to a collection of estimators

$\{\widehat{\theta}_{n;\psi} : \psi \in \Psi\}$ indexed by $\psi(\cdot)$, where each $\widehat{\theta}_{n;\psi}$ is an M-estimator of θ associated with the function $\psi(\cdot)$. The collection of M-estimators

$$\{\widehat{\theta}_{n;\psi} : \psi \in \Psi\}$$

is called functional M-process as in [16], where the parametric M-process was investigated. In this study, the case of intrinsic functional parameter is considered, namely θ_0 is a the common value that minimize for all the functions $\psi(\cdot)$ in a class Ψ, that is,

$$P\psi(\theta_0;\cdot) = 0, \text{ for all } \psi \in \Psi.$$

For convenience the centered process

$$\{\widehat{\theta}_{n;\psi} - \theta_0 : \psi \in \Psi\}$$

is considered. The main goal in the present paper is to provide a tool in order to construct estimators of $\widehat{\theta}_n$ using the contribution of the whole class Ψ rather than only one function from this class. More precisely, let T be a real-valued regular functional defined on the space $\ell^\infty(\Psi)$ of bounded functions defined on Ψ. By applying the functional T to a functional M-process, new kind of estimators can be defined which combine all the M-estimators in the class. This functional estimator of θ_0 is defined, in a natural way, by

$$\widetilde{\theta}_n := T(\{\widehat{\theta}_{n;\psi} : \psi \in \Psi\}) = T(\widehat{\theta}_{n,\bullet}).$$

The uniform convergence in probability on the class Ψ is established as well as the weak convergence of the processes $\{\widehat{\theta}_{n;\psi} - \theta_0 : \psi \in \Psi\}$ in the following section.

6.1. The large sample theory

Our large sample theory is a direct extension of the results given in [16]. We need the following definitions. For fixed $\gamma \in \mathbf{\Theta}$, the function

$$\varphi(\theta, \mathbb{P}_\gamma) = \mathbb{P}_\gamma\psi(\theta;\cdot),$$

as a map from $\mathbf{\Theta}$ to $\ell^\infty(\mathcal{H})$, is Fréchet differentiable with respect to the norm $\|\cdot\|$ at a point $\theta \in \mathbf{\Theta}$ if there is a bounded linear operator $\dot{\varphi}(\gamma, \mathbb{P}_\gamma)(\cdot)$ mapping from $(lin(\mathbf{\Theta}), \|\cdot\|)$, where $lin(\mathbf{\Theta})$ denote the linear span of $\mathbf{\Theta}$ (all linear combinations of elements in $\mathbf{\Theta}$), to $(\ell^\infty(\mathcal{H}), \|\cdot\|_\mathcal{H})$ such that

$$\|\varphi(\theta, \mathbb{P}_\gamma) - \varphi(\gamma, \mathbb{P}_\gamma) - \dot{\varphi}(\gamma, \mathbb{P}_\gamma)(\theta - \gamma)\|_\mathcal{H} = o(\|\theta - \gamma\|).$$

Denote the operator $\dot{\varphi}(\gamma, \mathbb{P}_\gamma)(\cdot)$ by $\dot{\varphi}(\gamma)$:

$$\dot{\varphi}(\gamma) := \dot{\varphi}(\gamma, \mathbb{P}_\gamma).$$

Suppose that for a fixed $\theta \in \mathbf{\Theta}$, the operator $\psi(\theta;\cdot)$ is bounded in the sense that

$$\|\mathbb{P}_\gamma\psi(\theta;\cdot)\|_\mathcal{H} < \infty$$

for all $\gamma \in \mathbf{\Theta}$. Thus for a fixed $\theta \in \mathbf{\Theta}$, the probability measure $\mathbb{P}_\gamma$ induces a mapping $\gamma \mapsto \mathbb{P}_\gamma\psi(\theta;\cdot)$ from $\mathbf{\Theta}$ to $\ell^\infty(\mathcal{H})$. The map $\mathbb{P}_\gamma\psi(\theta;\cdot)$, as a function of γ, is Fréchet differentiable with respect to the norm $\|\cdot\|$ at a point $\gamma \in \mathbf{\Theta}$ if there is a linear operator $\dot{P}_\gamma(\cdot)$ such that $\dot{P}_\gamma(\cdot)\psi(\theta;\cdot)$ is bounded and

$$\|\mathbb{P}_\gamma\psi(\theta;\cdot) - \mathbb{P}_\theta\psi(\theta;\cdot) - \dot{P}_\theta(\gamma - \theta)\psi(\theta;\cdot)\|_\mathcal{H} = o(\|\gamma - \theta\|).$$

7. 6.2. Weak convergence

We need the following assumptions for the weak convergence theorem.

A.1 For all $\gamma \in \mathbf{\Theta}$, $\mathbb{P}_\gamma \psi(\gamma; \cdot) = 0$ in $\ell^\infty(\mathcal{H})$ independently of ψ.

A.2 As $n \to \infty$, for any decreasing $\delta_n \downarrow 0$, the stochastic equicontinuity condition

$$\sup_{\psi \in \mathcal{F}} \sup_{\gamma} \{ \| \mathbb{G}_n(\psi(\gamma; \cdot) - \psi(\theta_0; \cdot)) \|_{\mathcal{H}} : \|\gamma - \theta_0\| \leq \delta_n \} = o_{\mathbb{P}^*}(1)$$

holds at the point θ_0, where $\mathbb{G}_n$ is the empirical process indexed by $\mathcal{F}$.

A.3 At the point θ_0,

$$\{\mathbb{G}_n \psi(\theta_0; \cdot) : \psi \in \mathcal{F}\} \rightsquigarrow \{\mathbb{Z}_0(\psi) : \psi \in \mathcal{F}\} \text{ in } \ell^\infty(\mathcal{H}),$$

where $\rightsquigarrow$ indicates weak convergence in $\ell^\infty(\mathcal{H})$ to a tight Borel measurable random element $\mathbb{Z}_0$.

A.4 For a fixed $\theta \in \mathbf{\Theta}$, the operator $\mathbb{P}_\gamma \psi(\theta; \cdot)$ as a function of γ is Fréchet differentiable with respect to the norm $\|\cdot\|$ at γ. Furthermore, the function $\theta \to \mathbb{P}_\gamma \psi(\theta; \cdot)$ from $\mathbf{\Theta}$ to $\ell^\infty(\mathcal{H})$ is Fréchet differentiable with respect to the norm $\|\cdot\|$ for all $\psi \in \mathcal{F}$. The operator $\dot{\varphi}(\gamma)$ is continuous as a function of γ in the following sense

$$\sup_{\psi \in \mathcal{F}} \|\dot{\varphi}(\gamma) - \dot{\varphi}(\theta)\| := \sup_{\psi \in \mathcal{F}} \sup_{\|a\| \leq 1} \|\dot{\varphi}(\gamma)(a) - \dot{\varphi}(\theta)(a)\|_{\mathcal{H}} \to 0, \text{ as } \|\gamma - \theta\| \to 0.$$

A.5 For every fixed $\gamma \in \mathbf{\Theta}$ and all $\psi \in \mathcal{F}$, the operator $\dot{\varphi}(\gamma)$ from $(\overline{lin(\mathbf{\Theta})}, \|\cdot\|)$, where $\overline{lin(\mathbf{\Theta})}$ the closure of $lin(\mathbf{\Theta})$, to $(\ell^\infty(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ has a bounded inverse $\dot{\varphi}^{-1}(\gamma)$ on a fixed subspace $\overline{\mathcal{R}(\dot{\varphi})} \subset \ell^\infty(\mathcal{H})$. Furthermore $\dot{\varphi}^{-1}(\gamma)$ as an operator sequence converges to $\dot{\varphi}^{-1}(\theta_0)$ as $\|\gamma - \theta_0\| \to 0$:

$$\sup_{\psi \in \mathcal{F}} \|\dot{\varphi}^{-1}(\gamma)(f) - \dot{\varphi}^{-1}(\theta_0)(f)\| \to 0,$$

for all $f \in \overline{\mathcal{R}(\dot{\varphi})}$.

Let us state our main result of this section.

Theorem 4. Let $\|\widehat{\theta}_{n,\psi} - \theta_0\| \to_{\mathbb{P}^*} 0$, for all $\psi \in \Psi$, be a sequence of consistent Z-estimator. Assume A.1 through A.5. Then we have

$$\{n^{1/2}(\widehat{\theta}_{n;\psi} - \theta_0) : \psi \in \Psi\} \rightsquigarrow \{\dot{\varphi}^{-1}(\theta_0)(\mathbb{Z}_0(\psi)) : \psi \in \Psi\},$$

where $\dot{\varphi}^{-1}(\theta_0)(\mathbb{Z}_0(\psi))$ is a centred Gaussian processes with covariance

$$\sigma(k, k') := \dot{\varphi}_k^{-1}(\theta_0)\dot{\varphi}_{k'}^{-1}(\theta_0) \int \psi_k(\theta_0; \cdot)\psi_{k'}(\theta_0; \cdot) dP,$$

for $\psi_k(\theta_0; \cdot)$ and $\psi_{k'}(\theta_0; \cdot)$ in Ψ.

An application of the continuous mapping theorem gives the following corollary.

Corollary 7.1. *Under the condition of Theorem 4, we have*

$$\max_{\psi \in \Psi_{\theta_0}} \{n^{1/2}(\widehat{\theta}_{n;\psi} - \theta_0) : \psi \in \Psi\} \rightsquigarrow \max_{\psi \in \Psi_{\theta_0}} \{\dot{\varphi}^{-1}(\theta_0)(\mathbb{Z}_0(\psi)) : \psi \in \Psi\},$$

and

$$\inf_{\psi \in \Psi_{\theta_0}} \{n^{1/2}(\widehat{\theta}_{n;\psi} - \theta_0) : \psi \in \Psi\} \rightsquigarrow \inf_{\psi \in \Psi_{\theta_0}} \{\dot{\varphi}^{-1}(\theta_0)(\mathbb{Z}_0(\psi)) : \psi \in \Psi\}.$$

8. Mathematical developpments

This section is devoted to the proofs of our results. The previously presented notation continues to be used in the following.

Proof of Theorem 1.

The proof is similar to that of Lemma 1 [40]. Since θ_0 is the unique solution to

$$\Psi(\theta, \alpha, \eta_0(\cdot; \theta)) = 0,$$

this implies that for any fixed $\epsilon > 0$, there exists a $\delta > 0$ such that

$$P\left[|\widehat{\alpha}_\phi(\theta) - \theta_0| > \epsilon\right] \leq P\left[|\Psi(\theta, \widehat{\alpha}_\phi(\theta), \eta_0(\cdot; \widehat{\alpha}_\phi(\theta)))| > \delta\right].$$

If we can prove

$$|\Psi(\theta, \widehat{\alpha}_\phi(\theta), \eta_0(\cdot; \widehat{\alpha}_\phi(\theta)))| \to_{\mathbb{P}^*} 0,$$

then the consistency of $\widehat{\alpha}_\phi(\theta)$ will follow immediately. To do this, first note that since

$$||\hat{\eta}_n - \eta_0|| = o_{\mathbb{P}^*}(1),$$

there exists a sequence $\{\delta_n\} \downarrow 0$ such that

$$||\hat{\eta}_n - \eta_0|| \leq \delta_n,$$

with probability tending to one. Hence taking $\eta = \hat{\eta}_n$ in equation (3.1), we have the following inequalities:

$$\begin{aligned}
&|\Psi(\theta, \widehat{\alpha}_\phi(\theta), \eta_0(\cdot; \widehat{\alpha}_\phi(\theta)))| \\
&\quad \leq |\Psi_n(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta)))| \\
&\qquad + |\Psi(\theta, \widehat{\alpha}_\phi(\theta), \eta_0(\cdot; \widehat{\alpha}_\phi(\theta))) - \Psi_n(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta)))| \\
&\quad \leq |\Psi_n(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta)))| + o_{\mathbb{P}^*}\Big(1 + |\Psi_n(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta)))| \\
&\qquad\qquad + |\Psi(\theta, \widehat{\alpha}_\phi(\theta), \eta_0(\cdot; \widehat{\alpha}_\phi(\theta)))|\Big) \\
&\quad \leq o_{\mathbb{P}^*}(1) + o_{\mathbb{P}^*}\left(1 + o_{\mathbb{P}^*}(1) + |\Psi(\theta, \widehat{\alpha}_\phi(\theta), \eta_0(\cdot; \widehat{\alpha}_\phi(\theta)))|\right),
\end{aligned}$$

which implies

$$|\Psi(\theta, \widehat{\alpha}_\phi(\theta), \eta_0(\cdot; \widehat{\alpha}_\phi(\theta)))| = o_{\mathbb{P}^*}(1).$$

So we have proved the consistency of pseudo Z-estimators $\widehat{\alpha}_\phi(\theta)$. □

Proof of Theorem 2.

We first show a result that we will use in the proof: under Conditions (A.1) and (A.2),

$$n^{1/2}\,|\Psi(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot, \widehat{\alpha}_\phi(\theta)))| = o_{\mathbb{P}^*}(1). \tag{8.1}$$

By Condition (A.1), we have the following inequality:

$$\begin{aligned}
&n^{1/2}\,|(\Psi_n - \Psi)(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta))) - (\Psi_n - \Psi)(\theta, \theta_0, \eta_0(\cdot; \theta_0))| \\
&\quad = o_{\mathbb{P}^*}(1) + o_{\mathbb{P}^*}\left(n^{1/2}\,|\Psi_n(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta)))|\right) \\
&\quad + o_{\mathbb{P}^*}\left(n^{1/2}\,|\Psi(\theta, \widehat{\alpha}_\phi(\theta), \hat{\eta}_n(\cdot; \widehat{\alpha}_\phi(\theta)))|\right).
\end{aligned}$$

By the triangle inequality

$$-|a| + |b| - |c| \leq |a - b - c|$$

and the fact that

$$\Psi(\theta, \theta_0, \eta_0(\cdot; \theta_0)) = 0,$$

we readily infer

$$\begin{aligned} & n^{1/2}\left|\Psi(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right| - n^{1/2}\left|\Psi_n(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right| \\ & \qquad -n^{1/2}\left|\Psi_n(\theta_0,\eta_0(\cdot;\theta_0))\right| \\ & \quad \leq \; n^{1/2}\left|(\Psi_n-\Psi)(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta))) - (\Psi_n-\Psi)(\theta_0,\eta_0(\cdot;\theta_0))\right| \\ & \quad = \; o_{\mathbb{P}^*}(1) + o_{\mathbb{P}^*}\left(n^{1/2}\left|\Psi_n(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right|\right) \\ & \qquad\quad + o_{\mathbb{P}^*}\left(n^{1/2}\left|\Psi(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right|\right), \end{aligned}$$

which implies

$$\begin{aligned} & n^{1/2}\left|\Psi(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right|\left[1-o_{\mathbb{P}^*}(1)\right] \\ & \quad \leq \; o_{\mathbb{P}^*}(1) + n^{1/2}\left|\Psi_n(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right|\left[1+o_{\mathbb{P}^*}(1)\right] \\ & \qquad + n^{1/2}\left|\Psi_n(\theta_0,\eta_0(\cdot;\theta_0))\right| \\ & \quad = \; o_{\mathbb{P}^*}(1) + o_{\mathbb{P}^*}(1) + o_{\mathbb{P}^*}(1). \end{aligned}$$

Hence (8.1) holds. We then show the root-n consistency of $\widehat{\alpha}_\phi(\theta)$. Since

$$|\widehat{\alpha}_\phi(\theta)-\theta_0| = o_{\mathbb{P}^*}(1),$$

and

$$\|\hat{\eta}_n-\eta_0\| = o_{\mathbb{P}^*}(n^{-\beta}),$$

with $\beta>0$, there exists a sequence $\{\delta_n\}\downarrow 0$ and $c>0$ such that

$$|\widehat{\alpha}_\phi(\theta)-\theta_0| \leq \delta_n$$

and

$$\|\hat{\eta}_n-\eta_0\| \leq cn^{-\beta},$$

with probability approaching one. Hence taking $(\alpha,\eta)=(\widehat{\alpha}_\phi(\theta),\hat{\eta}_n)$ in the smoothness condition (3.4):

$$\begin{aligned} & \Big| n^{1/2}\left\{\Psi(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta))) - \Psi(\theta_0,\eta_0(\cdot;\theta_0))\right\} \\ & \qquad - n^{1/2}\left\{\dot{\Psi}_1(\theta,\theta_0,\eta_0(\cdot;\theta_0)) + \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]\right\}(\widehat{\alpha}_\phi(\theta)-\theta_0) \\ & \qquad - n^{1/2}\dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[(\hat{\eta}_n-\eta_0)(\cdot;\theta_0)]\Big| \\ & \quad = \; o_{\mathbb{P}^*}\left(n^{1/2}|\widehat{\alpha}_\phi(\theta)-\theta_0|\right) + o_{\mathbb{P}^*}\left(n^{1/2}\|\hat{\eta}_n-\eta_0\|^{\xi}\right) \\ & \quad = \; o_{\mathbb{P}^*}\left(1+n^{1/2}|\widehat{\alpha}_\phi(\theta)-\theta_0|\right), \end{aligned} \tag{8.2}$$

since

$$n^{1/2}o_{\mathbb{P}^*}(\|\hat{\eta}_n-\eta_0\|^{\xi}) = o_{\mathbb{P}^*}(1)$$

by $\xi\beta>1/2$. Same result can be obtained by using the smoothness condition (3.3) for $\beta=1/2$. By equation (8.1), the fact that $\Psi(\theta,\theta_0,\eta_0(\cdot;\theta_0))=0$, and the triangle inequality

$$-|a|+|b|-|c| \leq |a-b-c|,$$

equation (8.2) implies

$$\begin{aligned} & -o_{\mathbb{P}^*}(1) + \left|n^{1/2}\left\{\dot{\Psi}_1(\theta,\theta_0,\eta_0(\cdot;\theta_0)) + \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]\right\}(\widehat{\alpha}_\phi(\theta)-\theta_0)\right| \\ & \qquad - \left|n^{1/2}\dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[(\hat{\eta}_n-\eta_0)(\cdot;\theta_0)]\right| \\ & \quad \leq \; o_{\mathbb{P}^*}\left(1+n^{1/2}\left|\widehat{\alpha}_\phi(\theta)-\theta_0\right|\right). \end{aligned} \tag{8.3}$$

Since the $d \times d$ matrix $\dot{\Psi}_1(\theta,\theta_0,\eta_0(\cdot;\theta_0)) + \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]$ is nonsingular, there exist a constant $c_1 > 0$ such that

$$\left|\left\{\dot{\Psi}_1(\theta,\theta_0,\eta_0(\cdot;\theta_0)) + \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]\right\}(\alpha-\theta_0)\right| \geq c_1|\alpha-\theta_0|$$

for

$$|\alpha-\theta_0| \to 0.$$

On the other hand, by Condition (A.4), combination with inequality (8.3) yields

$$\begin{aligned}
o_{\mathbb{P}^*}(1) \quad \geq \quad & \left|n^{1/2}\left\{\dot{\Psi}_1(\theta,\theta_0,\eta_0(\cdot;\theta_0)) + \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]\right\}(\widehat{\alpha}_\phi(\theta)-\theta_0)\right| \\
& - \left|n^{1/2}\dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[(\hat{\eta}_n-\eta_0)(\cdot;\theta_0)]\right| \\
& - o_{\mathbb{P}^*}\left(1+n^{1/2}\left|\widehat{\alpha}_\phi(\theta)-\theta_0\right|\right) \\
\geq \quad & c_1 n^{1/2}\left|\widehat{\alpha}_\phi(\theta)-\theta_0\right| - o_{\mathbb{P}^*}(1) - o_{\mathbb{P}^*}\left(1+n^{1/2}\left|\widehat{\alpha}_\phi(\theta)-\theta_0\right|\right) \\
= \quad & \left\{o_{\mathbb{P}^*}(1) - o_{\mathbb{P}^*}(1)\right\} n^{1/2}\left|\widehat{\alpha}_\phi(\theta)-\theta_0\right| - o_{\mathbb{P}^*}(1).
\end{aligned}$$

Hence the sequence $n^{1/2}\left|\widehat{\alpha}_\phi(\theta)-\theta_0\right|$ must be bounded in outer probability. Now we are ready to prove equation (3.5). Because

$$\begin{aligned}
& n^{1/2}\left[\Psi(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta))) - \Psi(\theta_0,\eta_0(\cdot;\theta_0))\right] \\
& \quad = n^{1/2}\left[\Psi(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta))) - \Psi_n(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right. \\
& \qquad \left. + \Psi_n(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta))) - \Psi(\theta_0,\eta_0(\cdot;\theta_0))\right] \\
& \quad = n^{1/2}(\Psi-\Psi_n)(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta))) + o_{\mathbb{P}^*}(1) - 0 \\
& \quad = -\,n^{1/2}(\Psi_n-\Psi)(\theta,\theta_0,\eta_0(\cdot;\theta_0)) \pm o_{\mathbb{P}^*}\Big(1+n^{1/2}\left|\Psi_n(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right| \\
& \qquad\qquad + n^{1/2}\left|\Psi(\theta,\widehat{\alpha}_\phi(\theta),\hat{\eta}_n(\cdot;\widehat{\alpha}_\phi(\theta)))\right|\Big) \quad \text{(by Condition (A.1))}
\end{aligned}$$

$$= -\,n^{1/2}(\Psi_n-\Psi)(\theta,\theta_0,\eta_0(\cdot;\theta_0)) \pm o_{\mathbb{P}^*}(1), \quad \text{(by equation (8.1))}, \tag{8.4}$$

after replacing equation (8.4) into the first term in the first line of equation (8.2) we obtain

$$\begin{aligned}
& \Big|-n^{1/2}(\Psi_n-\Psi)(\theta,\theta_0,\eta_0(\cdot;\theta_0)) \pm o_{\mathbb{P}^*}(1) - n^{1/2}\Big\{\dot{\Psi}_1(\theta,\theta_0,\eta_0(\cdot;\theta_0)) \\
& \qquad + \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]\Big\}(\widehat{\alpha}_\phi(\theta)-\theta_0) \\
& \qquad - n^{1/2}\dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[(\hat{\eta}_n-\eta_0)(\cdot;\theta_0)]\Big| \\
& \quad = o_{\mathbb{P}^*}\left(1+n^{1/2}\left|\widehat{\alpha}_\phi(\theta)-\theta_0\right|\right) \\
& \quad = o_{\mathbb{P}^*}(1),
\end{aligned}$$

which implies

$$\begin{aligned}
n^{1/2}(\widehat{\alpha}_\phi(\theta)-\theta_0) \; = \; & \Big\{-\dot{\Psi}_1(\theta_0,\eta_0(\cdot;\theta_0)) - \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[\dot{\eta}_0(\cdot;\theta_0)]\Big\}^{-1} \\
& \times n^{1/2}\Big\{(\Psi_n-\Psi)(\theta,\theta_0,\eta_0(\cdot;\theta_0)) \\
& \qquad + \dot{\Psi}_2(\theta,\theta_0,\eta_0(\cdot;\theta_0))[(\hat{\eta}_n-\eta_0)(\cdot;\theta_0)]\Big\} + o_{\mathbb{P}^*}(1).
\end{aligned}$$

Hence the proof is complete. □

Remark 8.1. Note that the proof techniques of Theorem 3 are largely inspired from that of [21] and changes have been made in order to adapt them to our purpose.

Proof of Theorem 3

Keep in mind the following definitions

$$\mathbb{G}_n := \sqrt{n}(\mathbb{P}_n - \mathbb{P}_{\theta_0,\eta_0})$$

and

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

In view of the fact that

$$\mathbb{P}_{\theta_0,\eta_0}\frac{\partial}{\partial\alpha}\widetilde{m}(\theta,\theta_0,\eta_0) = 0,$$

then a little calculation shows that

$$\begin{aligned}
&\mathbb{G}_n^*\widetilde{m}(\theta,\theta_0,\eta_0) + \mathbb{G}_n\widetilde{m}(\theta,\theta_0,\eta_0)\\
&\qquad +\sqrt{n}\mathbb{P}_{\theta_0,\eta_0}\left[\widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*) - \widetilde{m}(\theta,\theta_0,\eta_0)\right]\\
&= \quad \mathbb{G}_n^*\left[\widetilde{m}(\theta,\theta_0,\eta_0) - \widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*)\right]\\
&\qquad +\mathbb{G}_n\left[\widetilde{m}(\theta,\theta_0,\eta_0) - \widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*)\right]\\
&\qquad +\sqrt{n}\mathbb{P}_n^*\widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*).
\end{aligned}$$

Consequently, we have following inequality

$$\begin{aligned}
&\left\|\sqrt{n}\mathbb{P}_{\theta_0,\eta_0}\left[\widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*) - \widetilde{m}(\theta,\theta_0,\eta_0)\right]\right\|\\
&\quad\leq \quad \|\mathbb{G}_n^*\widetilde{m}(\theta,\theta_0,\eta_0)\| + \|\mathbb{G}_n\widetilde{m}(\theta,\theta_0,\eta_0)\|\\
&\qquad + \left\|\mathbb{G}_n^*\left[\widetilde{m}(\theta,\theta_0,\eta_0) - \widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*)\right]\right\|\\
&\qquad + \left\|\mathbb{G}_n\left[\widetilde{m}(\theta,\theta_0,\eta_0) - \widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*)\right]\right\|\\
&\qquad + \left\|\sqrt{n}\mathbb{P}_n^*\widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*)\right\|\\
&\quad := \quad G_1 + G_2 + G_3 + G_4 + G_5. \qquad (8.5)
\end{aligned}$$

According to Theorem 2.2 in [43], under condition (A.9), we have

$$G_1 = O_{\mathbb{P}_W}(1)$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability. In view of the CLT, we have

$$G_2 = O_{\mathbb{P}_{\theta_0,\eta_0}}(1).$$

By condition (A.10), we have

$$\|\widehat{\eta}_n^* - \eta_0\| = o_{\mathbb{P}_W}(1)$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability. This implies that

$$G_3 = o_{\mathbb{P}_W}(1)$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability, by using Lemma 3 in [21]. Similarly, we have

$$G_4 = o_{\mathbb{P}_W}(1)$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability. Finally we have,

$$G_4 = o_{\mathbb{P}_W}(1)$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability, by (5.3). Then we have

$$\left\|\sqrt{n}\mathbb{P}_{\theta_0,\eta_0}\left[\widetilde{m}(\theta,\widehat{\alpha}_\phi^*(\theta),\widehat{\eta}_n^*) - \widetilde{m}(\theta,\theta_0,\eta_0)\right]\right\| \leq O_{\mathbb{P}_W}(1) + O_{\mathbb{P}_{\theta_0,\eta_0}}(1) \qquad (8.6)$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability.

On the other hand, by a Taylor series expansion, we can write

$$\begin{aligned}
&\mathbb{P}_{\theta_0,\eta_0}\left[\widetilde{m}(\theta,\alpha,\widehat{\eta}_n) - \widetilde{m}(\theta,\theta_0,\eta_0)\right]\\
&\quad = \quad -(\alpha-\theta_0)^\top\Gamma_1 + O\left(\|\alpha-\theta_0\|^2 \vee n^{-1/2}\right). \qquad (8.7)
\end{aligned}$$

Clearly it is straightforward to combine (8.7) with (8.6), to infer the following

$$\begin{aligned}\sqrt{n}\left\|\Gamma_1\|\widehat{\alpha}^*_\phi(\theta)-\theta_0\|\right\| &\leq O^o_{\mathbb{P}_W}(1)+O^o_{\mathbb{P}_{\theta_0,\eta_0}}(1)\\ &+O^o_{\mathbb{P}_W}\left(\sqrt{n}\|\widehat{\alpha}^*_\phi(\theta)-\theta_0\|^2\right)\end{aligned} \tag{8.8}$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability. By considering again the consistency of $\widehat{\alpha}^*_\phi(\theta)$ and condition (A.2) and (A.6) and making use (8.8) to complete the proof of (5.6).
We next prove (5.7). Introduce

$$\begin{aligned}H_1 &:= -\mathbb{G}^*_n\left[\widetilde{m}(\theta,\widehat{\alpha}^*_\phi(\theta),\widehat{\eta}^*_n)-\widetilde{m}(\theta,\theta_0,\eta_0)\right],\\ H_2 &:= \mathbb{G}_n\left[\widetilde{m}(\theta,\widehat{\alpha}_\phi(\theta),\widehat{\eta}_n)-\widetilde{m}(\theta,\theta_0,\eta_0)\right],\\ H_3 &:= -\mathbb{G}_n\left[\widetilde{m}(\theta,\widehat{\alpha}^*_\phi(\theta),\widehat{\eta}^*_n)-\widetilde{m}(\theta,\theta_0,\eta_0)\right],\\ H_4 &:= \sqrt{n}\mathbb{P}^*_n\widetilde{m}(\theta,\widehat{\alpha}^*_\phi(\theta),\widehat{\eta}^*_n)-\sqrt{n}\mathbb{P}_n\widetilde{m}(\theta,\widehat{\alpha}_\phi(\theta),\widehat{\eta}_n).\end{aligned}$$

By some algebra, we obtain

$$\sqrt{n}\mathbb{P}_{\theta_0,\eta_0}\left(\widetilde{m}(\theta,\widehat{\alpha}^*_\phi(\theta),\widehat{\eta}^*_n)-\widetilde{m}(\theta,\widehat{\alpha}_\phi(\theta),\widehat{\eta}_n)\right)+\mathbb{G}^*_n\widetilde{m}(\theta,\theta_0,\eta_0)=\sum_{j=1}^4 H_j.$$

Using similar arguments to that of [21], to obtain

$$\begin{aligned}&\sqrt{n}\mathbb{P}_{\theta_0,\eta_0}\left(\widetilde{m}(\theta,\widehat{\alpha}^*_\phi(\theta),\widehat{\eta}^*_n)-\widetilde{m}(\theta,\widehat{\alpha}_\phi(\theta),\widehat{\eta}_n)\right)\\ &= -\mathbb{G}^*_n\widetilde{m}(\theta,\theta_0,\eta_0)+o_{\mathbb{P}_{\theta_0,\eta_0}}(1)\\ &\quad +o_{\mathbb{P}_W}(1)\end{aligned} \tag{8.9}$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability. To analyze the left hand side of (8.9), we rewrite it as

$$\begin{aligned}&\sqrt{n}\mathbb{P}_{\theta_0,\eta_0}\left[\widetilde{m}(\theta,\widehat{\alpha}^*_\phi(\theta),\widehat{\eta_n}^*)-\widetilde{m}(\theta,\theta_0,\eta_0)\right]\\ &-\sqrt{n}\mathbb{P}_{\theta_0,\eta_0}\left[\widetilde{m}(\theta,\widehat{\alpha}_\phi(\theta),\widehat{\eta}_n)-\widetilde{m}(\theta,\theta_0,\eta_0)\right].\end{aligned}$$

By a Taylor expansion, we obtain

$$\begin{aligned}&\sqrt{n}\mathbb{P}_{\theta_0,\eta_0}(m_{11}(\theta,\alpha,\eta)-m_{21}(\theta,\alpha,\eta)[H(\theta_0,\eta_0)])(\widehat{\alpha}^*_\phi(\theta)-\widehat{\alpha}_\phi(\theta))\\ &= \mathbb{G}^*_n\widetilde{m}(\theta,\theta_0,\eta_0)+o_{\mathbb{P}_{\theta_0,\eta_0}}(1)+o^o_{\mathbb{P}_W}(1)\\ &\quad +O_{\mathbb{P}_{\theta_0,\eta_0}}(n^{-1/2})+O^o_{\mathbb{P}_W}(n^{-1/2})\\ &= \mathbb{G}^*_n\widetilde{m}(\theta,\theta_0,\eta_0)+o_{\mathbb{P}_{\theta_0,\eta_0}}(1)+o^o_{\mathbb{P}_W}(1)\end{aligned} \tag{8.10}$$

in $\mathbb{P}_{\theta_0,\eta_0}$-probability. Keep in mind that, under condition (A.2) and (A.6), the matrix Γ_1 is nonsingular. Multiply both sides of (8.10) by Γ_1^{-1} to obtain (5.7). An application of [43, Lemma 4.6], under the bootstrap weight conditions, thus implies (5.8). Using [15, Theorem 3.2] and [50, Lemma 2.11], it easily follows that

$$\sup_{x\in\mathbb{R}^d}\left|\mathbb{P}_{\theta_0,\eta_0}(\sqrt{n}(\widehat{\alpha}_\phi(\theta)-\theta_0)\leq x)-\mathbb{P}(N(0,\Sigma)\leq x)\right|=o_{\mathbb{P}_{\theta_0,\eta_0}}(1). \tag{8.11}$$

By combining (5.8) and (8.11), we readily obtain the desired conclusion (5.9).

□

Proof of Theorem 4.

By the Fréchet differentiability of $\mathbb{P}_\gamma\psi(\theta;\cdot)$ at γ, for all $\psi\in\mathcal{F}$, we have

$$\mathbb{P}_\gamma\psi(\theta;\cdot)-\mathbb{P}_\theta\psi(\theta;\cdot)-\dot{P}_\theta(\gamma-\theta)\psi(\theta;\cdot)=o(\|\gamma-\theta\|).$$

Substituting $\widehat{\theta}_{n,\psi}$ for θ and θ_0 for γ, we obtain

$$\begin{aligned}&\dot{P}_{\widehat{\theta}_{n,\psi}}(\widehat{\theta}_{n,\psi}-\theta_0)\psi(\widehat{\theta}_{n,\psi};\cdot)\\ &\quad = \mathbb{P}_{\widehat{\theta}_{n,\psi}}\psi(\widehat{\theta}_{n,\psi};\cdot)-\mathbb{P}_{\theta_0}\psi(\widehat{\theta}_{n,\psi};\cdot)+o_{\mathbb{P}^*}(\|\widehat{\theta}_{n,\psi}-\theta_0\|)\\ &\quad = -\mathbb{P}_{\widehat{\theta}_{n,\psi}}\psi(\widehat{\theta}_{n,\psi};\cdot)+o_{\mathbb{P}^*}(\|\widehat{\theta}_{n,\psi}-\theta_0\|).\end{aligned}$$

Note that, for $\psi\in\mathcal{F}$,

$$\dot{\varphi}(\widehat{\theta}_{n,\psi})=\dot{\varphi}(\widehat{\theta}_{n,\psi},\mathbb{P}_{\widehat{\theta}_{n,\psi}}),$$

and by the identity (2.2) in Lemma 2.1 of [58] we have

$$\begin{aligned}&\dot{\varphi}(\widehat{\theta}_{n,\psi})(n^{1/2}(\widehat{\theta}_{n,\psi}-\theta_0))\\ &\quad = -n^{1/2}\dot{P}_{\widehat{\theta}_{n,\psi}}(\widehat{\theta}_{n,\psi}-\theta_0)\psi(\widehat{\theta}_{n,\psi};\cdot)\\ &\quad = -n^{1/2}\mathbb{P}_{\theta_0}\psi(\widehat{\theta}_{n,\psi};\cdot)+o_{\mathbb{P}^*}(n^{1/2}\|\widehat{\theta}_{n,\psi}-\theta_0\|)\\ &\quad = n^{1/2}\mathbb{P}_{\theta_0}\psi(\widehat{\theta}_{n,\psi};\cdot)+o_{\mathbb{P}^*}(1).\end{aligned}$$

The last equality follows from the consistency of $\widehat{\theta}_{n,\psi}$, A.1 through A.5 and Lemma 2.4 of [58]. Note that by A.5, for $\psi\in\mathcal{F}$, the operator sequence $\dot{\varphi}^{-1}(\gamma)$ converges to $\dot{\varphi}^{-1}(\theta_0)$ on as

$$\|\gamma-\theta_0\|\to 0.$$

Hence the Banach-Steinhaus Theorem and the consistency of $\widehat{\theta}_{n,\psi}$ imply that the operator norm of $\dot{\varphi}^{-1}(\widehat{\theta}_{n,\psi})$ is uniformly bounded in $\mathbb{P}^*$-probability when n is sufficiently large. It maps a term of $o_{\mathbb{P}^*}(1)$ in the $\|\cdot\|_{\mathcal{H}}$-norm into a term of $o_{\mathbb{P}^*}(1)$ in $\|\cdot\|$-norm:

$$\dot{\varphi}^{-1}(\widehat{\theta}_{n,\psi})(o_{\mathbb{P}^*}(1))=o_{\mathbb{P}^*}(1).$$

Making use of Lemma 2.3. of [58], this means that, for $\psi\in\mathcal{F}$,

$$\begin{aligned}&n^{1/2}(\widehat{\theta}_{n,\psi}-\theta_0)\\ &\quad = \dot{\varphi}^{-1}(\widehat{\theta}_{n,\psi})(n^{1/2}\mathbb{P}_{\theta_0}\psi(\widehat{\theta}_{n,\psi};\cdot)+o_{\mathbb{P}^*}(1))\\ &\quad = \dot{\varphi}^{-1}(\widehat{\theta}_{n,\psi})(\mathbb{G}_n\psi(\theta_0;\cdot)+o_{\mathbb{P}^*}(1))\\ &\quad = -\dot{\varphi}^{-1}(\widehat{\theta}_{n,\psi})(\mathbb{G}_n\psi(\theta_0;\cdot))+o_{\mathbb{P}^*}(1).\end{aligned}$$

By the triangle inequality and Lemma 2.2 of [58] we obtain

$$\dot{\varphi}^{-1}(\widehat{\theta}_{n,\psi})(\mathbb{G}_n\psi(\theta_0;\cdot))=\dot{\varphi}^{-1}(\theta_0)(\mathbb{G}_n\psi(\theta_0;\cdot))+o_{\mathbb{P}^*}(1).$$

By assumption (A.3), we have

$$\begin{aligned}&\{n^{1/2}(\widehat{\theta}_{n;\psi}-\theta_0):\psi\in\Psi\}\\ &\quad = \{\dot{\varphi}^{-1}(\theta_0)(\mathbb{G}_n\psi(\theta_0;\cdot)):\psi\in\Psi_{\theta_0}\}+o_{\mathbb{P}^*}(1)\\ &\qquad \rightsquigarrow \{\dot{\varphi}^{-1}(\theta_0)(\mathbb{Z}_0(\psi)):\psi\in\Psi_{\theta_0}\},\end{aligned}$$

where $\dot{\varphi}^{-1}(\theta_0)(\mathbb{Z}_0(\psi))$ is a centered Gaussian process with covariance given in Theorem 4, this complete the proof. □

Acknowledgement

The authors are indebted to the Editor-in-Chief and the referee for their very valuable comments and careful reading of the article.

References

1. S. Alvarez-Andrade and S. Bouzebda, *Strong approximations for weighted bootstrap of empirical and quantile processes with applications*, Stat. Methodol. **11** (2013), 36-52.
2. S. Alvarez-Andrade and S. Bouzebda, *On the local time of the weighted bootstrap and compound empirical processes*, Stoch. Anal. Appl. **33**(4) (2015), 609–629.
3. S. Alvarez-Andrade and S. Bouzebda, *Some selected topics for the bootstrap of the empirical and quantile processes*, Theory Stoch. Process. **24** (2019), no. 1, 19–48.
4. P. Barbe and P. Bertail, *The weighted bootstrap*, volume 98 of Lecture Notes in Statistics, Springer-Verlag, New York, 1995.
5. A. Basu, H. Shioya and C. Park, *Statistical inference*, volume 120 of Monographs on Statistics and Applied Probability, CRC Press, Boca Raton, FL, 2011. The minimum distance approach.
6. P. J. Bickel and Y. Ritov, *Efficient estimation in the errors in variables model*, Ann. Statist. **15**(2) (1987), 513–540.
7. P. J. Bickel, C. A. J. Klaassen, Y. Ritov and J. A. Wellner, *Efficient and adaptive estimation for semiparametric models*, Springer-Verlag, New York, 1998. Reprint of the 1993 original.
8. S. Bouzebda, *Bootstrap de l'estimateur de Hill: théorèmes limites*, Ann. I.S.U.P. **54**(1-2) (2010), 61–72.
9. S. Bouzebda, *On the strong approximation of bootstrapped empirical copula processes with applications*, Math. Methods Statist. **21**(3) (2012), 153–188.
10. S. Bouzebda and M. Cherfi, *General bootstrap for dual ϕ-divergence estimates*, J. Probab. Stat. **2012** (2012), 33 p.
11. S. Bouzebda and A. Keziou, *New estimates and tests of independence in semiparametric copula models*, Kybernetika (Prague) **46**(1) (2010), 178–201.
12. S. Bouzebda and N. Limnios, *On general bootstrap of empirical estimator of a semi-Markov Kernel with applications*, J. Multivariate Anal. **116** (2013), 52–62.
13. S. Bouzebda, Ch. Papamichail and N. Limnios, *On a multidimensional general bootstrap for empirical estimator of continuous-time semi-Markov kernels with applications*, J. Nonparametr. Stat. **30**(1) (2018), 49–86.
14. M. Broniatowski and A. Keziou, *Minimization of ϕ-divergences on sets of signed measures*, Studia Sci. Math. Hungar. **43**(4) (2006), 403–442.
15. M. Broniatowski and A. Keziou, *Parametric estimation and tests through divergences and the duality technique*, J. Multivariate Anal. **100**(1) (2009), 16–36.
16. F. Chebana, *Parametric estimation with a class of M-estimators*, Math. Methods Statist. **18**(3) (2009), 231–240.
17. J. Chen and J. D. Kalbfleisch, *Penalized minimum-distance estimates in finite mixture models*, Canad. J. Statist. **24**(2) (1996), 167–175.
18. J. H. Chen, *Optimal rate of convergence for finite mixture models*, Ann. Statist. **23**(1) (1995), 221–233.
19. X. Chen, O. Linton, and I. Van Keilegom, *Estimation of semiparametric models when the criterion function is not smooth*, Econometrica **71**(5) (2003), 1591–1608.
20. G. Cheng, *Moment Consistency of the Exchangeably Weighted Bootstrap for Semiparametric M-estimation*, Scand. J. Stat. **42**(3) (2015), 665–684.
21. G. Cheng and J. Z. Huang, *Bootstrap consistency for general semiparametric M-estimation*, Ann. Statist. **38**(5) (2010), 2884–2915.
22. M. Cherfi, *Dual ϕ-divergences estimation in normal models*, Arxiv preprint arXiv:1108.2999 (2011).
23. M. Cherfi, *Dual divergences estimation for censored survival data*, J. Statist. Plann. Inference **142**(7) (2012), 1746–1756.
24. N. Cressie and T. R. C. Read, *Multinomial goodness-of-fit tests*, J. Roy. Statist. Soc. Ser. B **46**(3) (1984), 440–464.
25. B. Efron, *Bootstrap methods: another look at the jackknife*, Ann. Statist. **7**(1) (1979), 1–26.
26. B. Efron and R. J. Tibshirani, *An introduction to the bootstrap* , Monographs on Statistics and Applied Probability, 57. Chapman and Hall, New York, 1993.
27. E. Giné and J. Zinn, *Bootstrapping general empirical measures*, Ann. Probab. **18**(2) (1990), 851–869.
28. P. Hall, *The bootstrap and Edgeworth expansion*, Springer Series in Statistics. Springer-Verlag, New York, 1992.
29. P. Hall, *On the nonparametric estimation of mixture proportions*, J. Roy. Statist. Soc. Ser. B **43**(2) (1981), 147–156.
30. A. Keziou and S. Leoni-Aubin, *On empirical likelihood for semiparametric two-sample density ratio models*, J. Statist. Plann. Inference **138**(4) (2008), 915–928.

31. M. R. Kosorok, *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York, 2008.
32. F. Liese and I. Vajda, *Convex statistical distances*, volume 95 of Teubner-Texte zur Mathematik [Teubner Texts in Mathematics]. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987. With German, French and Russian summaries.
33. F. Liese and I. Vajda, *On divergences and informations in statistics and information theory*, IEEE Trans. Inform. Theory **52**(10) (2006), 4394–4412.
34. A. Y. Lo, *A Bayesian method for weighted sampling*, Ann. Statist. **21**(4) (1993), 2138–2148.
35. L. F. James, *A study of a class of weighted bootstraps for censored data*, Ann. Statist. **25**(4) (1997), 1595–1621.
36. A. Janssen, *Resampling Student's t-type statistics*, Ann. Inst. Statist. Math. **57**(3) (2005), 507–529.
37. A. Janssen and T. Pauls, *How do bootstrap and permutation tests work?*, Ann. Statist. **31**(3) (2003), 768–806.
38. G. McLachlan and D. Peel, *Finite mixture models*, Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York, 2000.
39. S. A. Murphy and A. W. Van Der Vaart, *Likelihood inference in the errors-in-variables model*, J. Multivariate Anal. **59**(1) (1996), 81–108.
40. B. Nan and J. A. Wellner, *A general semiparametric Z-estimation approach for case-cohort studies*, Statistica Sinica **23** (2013), 1155–1180.
41. W. K. Newey and D. McFadden, *Large sample estimation and hypothesis testing*, In Handbook of econometrics, Vol. IV, volume 2 of Handbooks in Econom., pages 2111–2245. North-Holland, Amsterdam, 1994.
42. L. Pardo, *Statistical inference based on divergence measures*, volume 185 of Statistics: Textbooks and Monographs. Chapman & Hall/CRC, Boca Raton, FL., 2006.
43. J. Præstgaard and J. A. Wellner, *Exchangeably weighted bootstraps of the general empirical process*, Ann. Probab. **21**(4) (1993), 2053–2086.
44. D. M. Titterington, A. F. M. Smith and U. E. Makov, *Statistical analysis of finite mixture distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1985.
45. A. Toma and M. Broniatowski, *Dual divergence estimators and tests: robustness results*, J. Multivariate Anal. **102**(1) (2011), 20–36.
46. A. Toma and S. Leoni-Aubin, *Robust tests based on dual divergence estimators and saddlepoint approximations*, J. Multivariate Anal. **101**(5) (2010), 1143–1155.
47. A. van der Vaart, *Efficient maximum likelihood estimation in semiparametric mixture models*, Ann. Statist. **24**(2) (1996), 862–878.
48. A. van der Vaart, *Semiparametric statistics*, In Lectures on probability theory and statistics (Saint-Flour, 1999), volume 1781 of Lecture Notes in Math., pages 331–457. Springer, Berlin, 2002.
49. A. van der Vaart and J. A. Wellner, *Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes*, In High dimensional probability, II (Seattle, WA, 1999), volume 47 of Progr. Probab., pages 115–133, Birkhäuser Boston, Boston, MA, 2000.
50. A. W. van der Vaart, *Asymptotic statistics*, volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1998.
51. A. W. van der Vaart and J. A. Wellner, *Weak convergence and empirical processes*, Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.
52. W. R. van Zwet, *The Edgeworth expansion for linear combinations of uniform order statistics*, In Proceedings of the Second Prague Symposium on Asymptotic Statistics (Hradec Králové, 1978), pages 93–101. North-Holland, Amsterdam-New York, 1979.
53. J. A. Wellner and Y. Zhan, *Bootstrapping z-estimators*, Technical Report 308, July 1996.
54. C.-S. Weng, *On a second-order asymptotic property of the Bayesian bootstrap mean*, Ann. Statist. **17**(2) (1989), 705–710.
55. J. Wu and R. J. Karunamuni, *Efficient Hellinger distance estimates for semiparametric models*, J. Multivariate Anal. **107** (2012), 1–23.
56. J. Wu, *Minimum Hellinger distance estimation in semiparametric models*, ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)–University of Alberta (Canada), 2008.
57. C.F.J. Wu, *On the asymptotic property of the jackknife histogram*, Technical report, Dept. of Statistics, Univ. of Wisconsin, Madison, 1987.
58. Y. Zhan, *Central limit theorems for functional Z-estimators*, Statist. Sinica **12**(2) (2002), 609–634.
59. Z. G. Zheng and D. S. Tu, *Random weighting methods in regression models*, Sci. Sinica Ser. A **31**(12) (1988), 1442–1459.

Alliance Sorbonne Université, Université de Technologie de Compiègne,
L.M.A.C., Compiègne, France
E-mail address: salim.bouzebda@utc.fr

Département de Mathématiques, Faculté des Sciences Exactes et Informatique,
University of Chlef (Université Hassiba Benbouali)
E-mail address: mohamed.cherfi@gmail.com