

# POSTERIOR DISTRIBUTION OF NONDIFFERENTIABLE FUNCTIONS.<sup>1</sup>

TORU KITAGAWA<sup>2</sup>, JOSÉ-LUIS MONTIEL-OLEA<sup>3</sup> AND JONATHAN PAYNE<sup>4</sup>

This paper examines the asymptotic behavior of the posterior distribution of a possibly nondifferentiable function  $g(\theta)$ , where  $\theta$  is a finite-dimensional parameter of either a parametric or semiparametric model. The main assumption is that the distribution of a suitable estimator  $\hat{\theta}_n$ , its bootstrap approximation, and the Bayesian posterior for  $\theta$  all agree asymptotically.

It is shown that whenever  $g$  is Lipschitz, though not necessarily differentiable, the posterior distribution of  $g(\theta)$  and the bootstrap distribution of  $g(\hat{\theta}_n)$  coincide asymptotically. One implication is that Bayesians can interpret bootstrap inference for  $g(\theta)$  as approximately valid posterior inference in a large sample. Another implication—built on known results about bootstrap inconsistency—is that credible sets for a nondifferentiable parameter  $g(\theta)$  cannot be presumed to be approximately valid confidence sets (even when this relation holds true for  $\theta$ ).

KEYWORDS: Bootstrap, Bernstein-von Mises Theorem, Directional Differentiability, Posterior Inference.

## 1. INTRODUCTION

This paper studies the posterior distribution of a real-valued function  $g(\theta)$ , where  $\theta$  is a parameter of finite dimension in either a parametric or semiparametric model. We focus on transformations  $g(\theta)$  that are Lipschitz continuous but possibly nondifferentiable. Some stylized examples are:

$$|\theta|, \max\{0, \theta\}, \max\{\theta_1, \theta_2\}.$$

Parameters of the type considered in this paper arise in different applications in economics and statistics: the welfare level attained by an optimal treatment assignment rule in the treatment choice problem (Manski (2004)); the regression function in a

---

<sup>1</sup>We would like to thank Gary Chamberlain, Tim Cogley, Hiroaki Kaido, Ulrich Müller, Mikkel Plagborg-Møller, Andres Santos, Quang Vuong, and two anonymous referees for detailed comments and suggestions on a earlier draft of this paper. All errors remain our own. Financial support from the ESRC through the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001) and from the ERC through the ERC starting grant (grant number EPP-715940) is gratefully acknowledged. This draft: September 28, 2017.

<sup>2</sup>University College London, Department of Economics. Kyoto University, Department of Economics. E-mail: [t.kitagawa@ucl.ac.uk](mailto:t.kitagawa@ucl.ac.uk).

<sup>3</sup>Columbia University, Department of Economics. E-mail: [montiel.olea@gmail.com](mailto:montiel.olea@gmail.com).

<sup>4</sup>New York University, Department of Economics. E-mail: [jep459@nyu.edu](mailto:jep459@nyu.edu).

regression kink model with an unknown threshold (Hansen (2015)); the eigenvalues of a random symmetric matrix (Eaton and Tyler (1991)); and the value function of stochastic mathematical programs (Shapiro (1991)). The lower and upper bound of the identified set in a partially identified model are also examples of parameters that fall within the framework of this paper.<sup>1</sup>

The potential nondifferentiability of  $g(\cdot)$  poses different challenges to frequentist inference. For example, different forms of the bootstrap lose their consistency whenever differentiability is compromised; see Dümbgen (1993), Beran (1997), Andrews (2000), Fang and Santos (2015), and Hong and Li (2015). To our knowledge, the literature has not yet explored how the Bayesian posterior of  $g(\theta)$  relates to neither the sampling nor the bootstrap distribution of available plug-in estimators when  $g$  is allowed to be nondifferentiable.

This paper studies these relations in large samples. The main assumptions are that: (i) there is an estimator for  $\theta$ , denoted  $\hat{\theta}_n$ , which is  $\sqrt{n}$ -asymptotically distributed according to some random vector  $Z$  (not necessarily Gaussian), (ii) the bootstrap consistently estimates the asymptotic distribution of  $\hat{\theta}_n$  and (iii) the Bayesian posterior distribution of  $\theta$  coincides with the asymptotic distribution of  $\hat{\theta}_n$ ; i.e., the Bernstein-von Mises Theorem holds for  $\theta$ .<sup>2</sup>

This paper shows that—after appropriate centering and scaling—the posterior distribution of  $g(\theta)$  and the bootstrap distribution of  $g(\hat{\theta}_n)$  are asymptotically equivalent. This means that the bootstrap distribution of  $g(\hat{\theta}_n)$  contains, in large samples, the same information as the posterior distribution for  $g(\theta)$ .<sup>3</sup> Indisputably, these asymptotic relations are straightforward to deduce for (fully or directionally) differentiable functions. However, our main result shows that the asymptotic equivalence between the bootstrap and posterior distributions holds more broadly; highlighting that such a relation is better understood as a consequence of the continuous mapping

---

<sup>1</sup>For example, treatment effect bounds (Manski (1990), Balke and Pearl (1997)); bounds in auction models (Haile and Tamer (2003)); bounds for impulse-response functions (Giacomini and Kitagawa (2015), Gafarov, Meier, and Montiel Olea (2015)) and forecast-error variance decompositions (Faust (1998)) in Structural Vector Autoregressions.

<sup>2</sup>See, for example, DasGupta (2008), p. 291 for a Bernstein-von Mises theorem for regular parametric models where  $Z$  is Gaussian; Ghosal, Ghosh, and Samanta (1995), p. 2147-2150 for a Bernstein-von Mises theorem for a class of parametric models whose likelihood ratio process is not Locally Asymptotically Normal; and Castillo and Rousseau (2015), p. 2357 for a Bernstein-von Mises theorem for semiparametric models where an efficiency theory at rate  $\sqrt{n}$  is available.

<sup>3</sup>Other results in the literature concerning the relations between bootstrap and posterior inference have focused on the Bayesian interpretation of the bootstrap in finite samples, for example Rubin (1981), or on how the parametric bootstrap output can be used for efficient computation of the posterior, for example Efron (2012).

theorem, as opposed to differentiability and the delta method.

This result provides two useful insights. First, Bayesians can interpret bootstrap-based inference for  $g(\theta)$  as approximately valid posterior inference in a large sample. Thus, Bayesians can use bootstrap draws to conduct approximate posterior inference for  $g(\theta)$  (if computing  $\hat{\theta}_n$  is simpler than Markov Chain Monte Carlo (MCMC) sampling).

Second, we show that whenever nondifferentiability causes a bootstrap confidence set to cover  $g(\theta)$  less often than desired—which is known to happen even under mild departures from differentiability—a credible set based on the quantiles of the posterior will have distorted frequentist coverage as well. In the case where  $g(\cdot)$  only has directional derivatives, as in the pioneering work of [Hirano and Porter \(2012\)](#), the unfortunate frequentist properties of credible sets can be attributed to the fact that the posterior distribution of  $g(\theta)$  does not coincide with the asymptotic distribution of  $g(\hat{\theta}_n)$ .

The rest of this paper is organized as follows. Section 2 presents a formal statement of the main results. Section 3 presents an illustrative example: the absolute value transformation. Section 4 concludes. All the proofs are collected in the Appendix.

## 2. MAIN RESULTS

Let  $X^n = \{X_1, \dots, X_n\}$  be a sample of size  $n$  from the model  $f(X^n | \eta)$ , where  $\eta$  is a possibly infinite dimensional parameter taking values in some space  $S$ . We assume there is a *finite-dimensional* parameter of interest,  $\theta : S \rightarrow \Theta \subseteq \mathbb{R}^p$ , and some estimator  $\hat{\theta}_n$  of  $\theta$ . Let  $\theta_0$  denote the true parameter—that is,  $\theta_0 \equiv \theta(\eta_0)$  with data generated according to  $f(X^n | \eta_0)$ . Consider the following assumptions:

**ASSUMPTION 1** The function  $g : \mathbb{R}^p \rightarrow \mathbb{R}$  is Lipschitz continuous with constant  $c$ . That is;

$$|g(x) - g(y)| \leq c \|x - y\| \quad \forall x, y \in \mathbb{R}^p.$$

Assumption 1 implies—by means of the well-known Rademacher’s Theorem ([Evans and Garipey \(2015\)](#), p. 81)—that  $g$  is differentiable almost everywhere in  $\mathbb{R}^p$ . Thus, the functions considered in this paper allow only for mild departures from differentiability.<sup>4</sup> We have made Lipschitz continuity our starting point—as opposed to

---

<sup>4</sup>Moreover, we assume that  $g$  is defined everywhere in  $\mathbb{R}^p$  which rules out examples such as the ratio of means  $\theta_1/\theta_2$ ,  $\theta_2 \neq 0$  discussed in [Fieller \(1954\)](#) and weakly identified Instrumental Variables models.

some form of directional differentiability—to emphasize that the asymptotic relation between Bootstrap and Bayes inference does not hinge on delta-method considerations.

**ASSUMPTION 2** The sequence  $Z_n \equiv \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} Z$ .

Despite being high-level, there are well-known conditions for parametric or semi-parametric models under which Assumption 2 obtains (see, for example, [Newey and McFadden \(1994\)](#) p. 2146). The asymptotic distribution of  $Z_n$  is typically normal, but our main theorems does not exploit this feature (and thus, we have decided leave the distribution of  $Z$  unspecified).

In order to state the next assumption, we introduce additional notation. Define the set:

$$BL(1) \equiv \left\{ f : \mathbb{R}^p \rightarrow \mathbb{R} \mid \sup_{a \in \mathbb{R}^k} |f(a)| \leq 1 \text{ and } |f(a_1) - f(a_2)| \leq \|a_1 - a_2\| \quad \forall a_1, a_2 \right\}.$$

Let  $\phi_n^*$  and  $\psi_n^*$  be random variables whose distribution depends on the data  $X^n$ . The Bounded Lipschitz distance between the distributions induced by  $\phi_n^*$  and  $\psi_n^*$  (conditional on the data  $X^n$ ) is defined as:

$$\beta(\phi_n^*, \psi_n^*; X^n) \equiv \sup_{f \in BL(1)} \left| \mathbb{E}[f(\phi_n^*) | X^n] - \mathbb{E}[f(\psi_n^*) | X^n] \right|.$$

The random variables  $\phi_n^*$  and  $\psi_n^*$  are said to converge *in Bounded Lipschitz distance in probability* if  $\beta(\phi_n^*, \psi_n^*; X^n) \xrightarrow{P} 0$  as  $n \rightarrow \infty$ .<sup>5</sup>

Let  $P$  denote some prior for  $\theta$  and let  $\theta_n^{P*}$  denote the random variable with law equal to the posterior distribution of  $\theta$  in a sample of size  $n$ . Let  $\theta_n^{B*}$  denote the random variable with law equal to the bootstrap distribution of  $\hat{\theta}_n$  in a sample of size  $n$ .

**REMARK 1** In a parametric model there are different ways of bootstrapping the distribution of  $\hat{\theta}_n$ . One possibility is a *parametric bootstrap*, which consists in generating draws  $(x_1, \dots, x_n)$  from the model  $f(x_i; \hat{\theta}_n)$  followed by an evaluation of the ML estimator for each draw ([Van der Vaart \(2000\)](#) p. 328). Another possibility is the *multinomial bootstrap*, which generates draws  $(x_1, \dots, x_n)$  from its empirical distribution. Different options are also available in semiparametric models. We do not

<sup>5</sup>For a more detailed treatment of the bounded lipschitz metric over probability measures see the ‘ $\beta$ ’ metric defined in p. 394 of [Dudley \(2002\)](#).

take a stand on the specific bootstrap procedure used by the researcher as long as it is consistent.

The following assumption restricts the prior  $P$  for  $\theta$  and the bootstrap procedure for  $\hat{\theta}_n$ :

**ASSUMPTION 3** The centered and scaled random variables:

$$Z_n^{P*} \equiv \sqrt{n}(\theta_n^{P*} - \hat{\theta}_n) \quad \text{and} \quad Z_n^{B*} \equiv \sqrt{n}(\theta_n^{B*} - \hat{\theta}_n),$$

converge (in the Bounded Lipschitz distance in probability) to the asymptotic distribution of  $\hat{\theta}_n$ , denoted  $Z$ , which is independent of the data. That is,

$$\beta(Z_n^{P*}, Z; X^n) \xrightarrow{P} 0$$

and

$$\beta(Z_n^{B*}, Z; X^n) \xrightarrow{P} 0.$$

Sufficient conditions for Assumption 3 to hold are the consistency of the bootstrap for the distribution of  $\hat{\theta}_n$  (Horowitz (2001), Van der Vaart and Wellner (1996) Chapter 3.6, Van der Vaart (2000) p. 340) and the Bernstein-von Mises Theorem for  $\theta$  (see DasGupta (2008) for parametric versions and Castillo and Rousseau (2015) for semiparametric ones).<sup>6</sup>

The following theorem shows that under the first three assumptions, the Bayesian posterior for  $g(\theta)$  and the frequentist bootstrap distribution of  $g(\hat{\theta}_n)$  converge (after appropriate centering and scaling). Note that for any measurable function  $g(\cdot)$ , be it differentiable or not, the posterior distribution of  $g(\theta)$  can be defined as the *image measure* induced by the distribution of  $\theta_n^{P*}$  under the mapping  $g(\cdot)$ .

**THEOREM 1** *Suppose that Assumptions 1, 2 and 3 hold. Then,*

$$\beta(\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)), \sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)); X^n) \xrightarrow{P} 0.$$

---

<sup>6</sup>Note that the Bernstein-von Mises Theorem is oftentimes stated in terms of almost-sure convergence of the posterior to a Gaussian distribution (DasGupta (2008) p. 291) or possibly to a non-Gaussian limit (Ghosal et al. (1995)) in terms of the total variation distance. This mode of convergence (total variation metric) implies convergence in the bounded Lipschitz metric in probability. In this paper, all the results concerning the asymptotic behavior of the posterior are presented in terms of the Bounded-Lipschitz metric. This facilitates comparisons with the bootstrap.

That is, after centering and scaling, the posterior distribution  $g(\theta)$  and the bootstrap distribution of  $g(\hat{\theta}_n)$  are asymptotically close to each other in terms of the Bounded Lipschitz metric in probability.

PROOF: See Appendix A.1.

*Q.E.D.*

The intuition behind Theorem 1 is the following. The centered and scaled posterior and bootstrap distributions can be written as:

$$\begin{aligned}\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)) &= \sqrt{n}(g(\theta_0 + Z_n^{P*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n)), \\ \sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)) &= \sqrt{n}(g(\theta_0 + Z_n^{B*}/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))\end{aligned}$$

Since  $Z_n^{P*}$  and  $Z_n^{B*}$  both converge to a common limit  $Z$  and the function  $g$  is Lipschitz, then the centered and scaled posterior and bootstrap distributions (conditional on the data) can both be well approximated by:

$$\sqrt{n}(g(\theta_0 + Z/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$$

and so the desired convergence result obtains. As the proof of the theorem illustrates, the asymptotic relation between the Bootstrap and Bayes distributions is a consequence of a (Lipschitz) continuous mapping theorem, and not of the delta method.

**FAILURE OF BOOTSTRAP/BAYES INFERENCE:** Theorem 1 established the large-sample equivalence between the bootstrap distribution of  $g(\hat{\theta}_n)$  and the posterior distribution of  $g(\theta)$ . We now use this Theorem to make a concrete connection between the coverage of bootstrap-based confidence sets and the coverage of Bayesian credible sets based on the quantiles of the posterior.

We start by assuming that a nominal  $(1 - \alpha)$  bootstrap confidence set fails to cover  $g(\theta)$  at a point of nondifferentiability. Then, we show that a  $(1 - \alpha - \epsilon)$  credible set based on the quantiles of the posterior distribution of  $g(\theta)$  will also fail to cover  $g(\theta)$  for any  $\epsilon > 0$ .<sup>7</sup>

This result is not a direct corollary of Theorem 1 as there is some extra work needed to relate the quantiles of the bootstrap distribution of  $g(\hat{\theta}_n)$  and the quantiles

---

<sup>7</sup>The adjustment factor  $\epsilon$  is introduced because the the quantiles of both the bootstrap and the posterior remain random even in large samples.

of the posterior of  $g(\theta)$ . To establish such connection, we work under the following assumption:

**ASSUMPTION 4** There exists a function  $h_{\theta_0}(Z, X^n)$  such that:

- i)  $\beta(\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n)), h_{\theta_0}(Z, X^n); X^n) \xrightarrow{P} 0$ .
- ii) The cumulative distribution function of  $Y \equiv h_{\theta_0}(Z, X^n)$  conditional on  $X^n$ , denoted  $F_{\theta_0}(y|X^n)$ , is Lipschitz continuous in  $y$ —almost surely in  $X^n$  for every  $n$ —with a constant  $k$  that does not depend on  $X^n$ .

The first part of Assumption 4 simply requires the distribution of  $\sqrt{n}(g(\theta_n^{B*}) - g(\theta_n))$ , conditional on the data, to have a well defined limit (which is neither assumed nor guaranteed by Theorem 1).

**REMARK 2** Note that the first part of Assumption 4 is satisfied if  $g$  is directionally differentiable, i.e., there exists a continuous function  $g'_{\theta_0} : \mathbb{R}^p \rightarrow \mathbb{R}$  such that for any compact set  $K \subseteq \mathbb{R}^p$  and any sequence of positive numbers  $t_n \rightarrow 0$ :<sup>8</sup>

$$\sup_{h \in K} \left| t_n^{-1}(g(\theta_0 + t_n h) - g(\theta_0)) - g'_{\theta_0}(h) \right| \rightarrow 0.$$

Assumption 4 then holds with  $h_{\theta_0}(Z, X_n) = g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)$  by the delta method for directionally differentiable functions shown in Proposition 1 in [Dümbgen \(1993\)](#) and equation A.41 in Theorem A.1 in [Fang and Santos \(2015\)](#).<sup>9</sup>

The second part of Assumption 4 requires the limiting distribution of the bootstrap to be well-behaved enough at points of possible nondifferentiability. In particular, we

<sup>8</sup>Equivalently, one could say there is a continuous function  $g'_\theta : \mathbb{R}^k \rightarrow \mathbb{R}$  such that for any converging sequence  $h_n \rightarrow h$ :

$$\left| \sqrt{n} \left( g \left( \theta_0 + \frac{h_n}{\sqrt{n}} \right) - g(\theta_0) \right) - g'_\theta(h_n) \right| \rightarrow 0.$$

See p. 479 in [Shapiro \(1990\)](#). The continuous, not necessarily linear, function  $g'_\theta(\cdot)$  will be referred to as the (Hadamard) directional derivative of  $g$  at  $\theta_0$ .

<sup>9</sup>For the sake of completeness, Lemma 4 in Appendix A.3 shows that if Assumptions 1, 2, 3 hold and  $g$  is directionally differentiable (in the sense defined in Remark 2). Then,

$$\beta(\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)), g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n); X^n) \xrightarrow{P} 0$$

holds, where  $Z$  is as defined in Assumption 3 and  $Z_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ .

exploit the Lipschitz continuity of the limiting distribution to relate the bootstrap quantiles with nearby quantiles of the posterior distribution. A sufficient condition for the second part of Assumption 4 to hold is that the density of  $Y$ , conditional on the data, admits an upper bound independent of  $X^n$ . This will be the case in the illustrative example we consider.

We now present the main definitions that will be used in the statement of next Theorem.

SET-UP FOR THEOREM 2: Let  $q_\alpha^B(X^n)$  be defined as:

$$q_\alpha^B(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^{B^*}(g(\theta_n^{B^*}) \leq c \mid X^n) \geq \alpha\}.$$

The quantile based on the posterior distribution  $q_\alpha^P(X^n)$  is defined analogously. A nominal  $(1 - \alpha)\%$  two-sided confidence set for  $g(\theta)$  based on the bootstrap distribution  $g(\theta_n^{B^*})$  can be defined as follows:

$$(2.1) \quad CS_n^B(1 - \alpha) \equiv [q_{\alpha/2}^B(X^n), q_{1-\alpha/2}^B(X^n)].$$

This is a typical confidence set based on the percentile method of Efron, p. 327 in [Van der Vaart \(2000\)](#).

**DEFINITION** We say that the nominal  $(1 - \alpha)\%$  bootstrap confidence set fails to cover the parameter  $g(\theta)$  at  $\theta$  by at least  $d_\alpha\%$  ( $d_\alpha > 0$ ) if:

$$(2.2) \quad \limsup_{n \rightarrow \infty} \mathbb{P}_\theta \left( g(\theta) \in CS_n^B(1 - \alpha) \right) \leq 1 - \alpha - d_\alpha,$$

where  $\mathbb{P}_\theta$  refers to the distribution of  $X_i$  under parameter  $\theta$ .

The next theorem shows the coverage probability of the Bayesian credible set for  $g(\theta)$  in relation to the coverage probability of its bootstrap confidence set.

**THEOREM 2** *Suppose that the nominal  $(1 - \alpha)\%$  bootstrap confidence set fails to cover  $g(\theta)$  at  $\theta$  by at least  $d_\alpha\%$ . If Assumptions 1 to 4 hold then for any  $\epsilon > 0$ :*

$$\limsup_{n \rightarrow \infty} \mathbb{P}_\theta \left( g(\theta) \in [q_{(\alpha+\epsilon)/2}^P(X^n), q_{1-(\alpha+\epsilon)/2}^P(X^n)] \right) \leq 1 - \alpha - d_\alpha.$$



Thus, for any  $0 < \epsilon < d_\alpha$ , the nominal  $(1 - \alpha - \epsilon)\%$  credible set based on the quantiles of the posterior fails to cover  $g(\theta)$  at  $\theta$  by at least  $(d_\alpha - \epsilon)\%$ .

PROOF: See Appendix A.2.

*Q.E.D.*

The intuition behind the theorem is the following. For convenience, let  $\theta_n^*$  denote either the bootstrap or posterior random variable and let  $c_\beta^*(X^n)$  denote the  $\beta$ -critical value of  $g(\theta_n^*)$  defined by:

$$c_\beta^*(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n)) \leq c \mid X^n) \geq \beta\}.$$

We show that  $c_\beta^*(X^n)$  is asymptotically close to the conditional  $\beta$ -quantile of  $h_\theta(Z, X^n)$ , denoted by  $c_\beta(X^n)$ . More precisely, we show that for arbitrarily small  $0 < \epsilon < \beta$  and  $\delta > 0$ , the probability that  $c_\beta^*(X^n) \in [c_{\beta-\epsilon/2}(X^n), c_{\beta+\epsilon/2}(X^n)]$  is greater than  $1 - \delta$  for sufficiently large  $n$ . Note that this result cannot be obtained directly from the fact that the difference between  $c_\beta^*(X^n)$  and  $c_\beta(X^n)$  is  $o_p(1)$ .

Because under Assumptions 1 to 4 the critical values of both the bootstrap and posterior distributions are asymptotically close to the quantiles of  $h_\theta(Z, X^n)$ , we can show that for a fixed  $\epsilon > 0$  and sufficiently large  $n$ :

$$\mathbb{P}_\theta \left( g(\theta) \in CS_n^B(1 - \alpha) \right) = \mathbb{P}_\theta \left( g(\theta) \in \left[ q_{(\alpha+\epsilon)/2}^P(X^n), q_{1-(\alpha+\epsilon)/2}^P(X^n) \right] \right) - \delta.$$

It follows that when the  $(1 - \alpha)\%$ -bootstrap confidence set fails to cover the parameter  $g(\theta)$  at  $\theta$ , then so must the  $(1 - \alpha - \epsilon)\%$ -credible set.<sup>10</sup>

---

<sup>10</sup>It immediately follows that the reverse also applies. If the  $(1 - \alpha)\%$ -credible set fails to cover the parameter  $g(\theta)$  at  $\theta$ , then so must the  $(1 - \alpha - \epsilon)\%$ -bootstrap confidence set. Note that our approximation holds for any fixed  $\epsilon$ , but we cannot guarantee that our approximation holds if we take the limit.

POSTERIOR DISTRIBUTION OF  $g(\theta^{P*})$  UNDER DIRECTIONAL DIFFERENTIABILITY: In Theorem 2 we chose to remain agnostic about the specific form of  $h_\theta(Z, X^n)$  at points of nondifferentiability. Our theorem shows that any assumption about the specific form of nondifferentiability—such as the existence of directional derivatives—does not play any role in establishing the asymptotic relation between Bootstrap and Posterior inference.

There are some benefits, however, in being more explicit about the form in which differentiability is violated. For instance, if  $g(\cdot)$  is assumed to be directionally differentiable (as defined in Remark 2) the posterior  $g(\theta^{P*})$  can be characterized more explicitly.

Lemma 4 in Appendix A.3 shows that if Assumptions 1, 2, 3 hold and  $g$  is directionally differentiable in the sense defined in Remark 2, then:

$$\beta(\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n)), g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n); X^n) \xrightarrow{P} 0,$$

where  $Z$  is as defined in Assumption 2 and  $Z_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ .

The distribution  $g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)$  (which still depends on the sample size) provides a large-sample approximation to the distribution of  $g(\theta_n^{P*})$ . Our result shows that, in large samples, after centering around  $g(\hat{\theta}_n)$ , the data will only affect the posterior distribution through  $Z_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ .

The approximating distribution  $g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)$  has appeared in the literature before, see Proposition 1 in [Dümbgen \(1993\)](#) and equation A.41 in Theorem A.1 in [Fang and Santos \(2015\)](#). Thus, verifying the assumptions for any of these papers in combination with our Theorem 1 would suffice to derive the limiting distribution of  $g(\theta^{P*})$  under directional differentiability. In order to keep the exposition self-contained, we decided to present a simpler derivation of this law using our Lipschitz continuity assumption.

Note that if  $g'_{\theta_0}(\cdot)$  is linear (which is the case if  $g$  is fully differentiable), then the derivative can be characterized by a vector  $g'_{\theta_0}$  and  $\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n))$  converges to:

$$g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n) = g'_{\theta_0}(Z),$$

where  $(g'_{\theta_0})^T$  denotes the transpose of  $g'_{\theta_0}$ . This is the same limit as one would get from applying the delta method to  $g(\hat{\theta}_n)$ . Thus, under full differentiability, the

posterior distribution of  $g(\theta)$  can be approximated as:

$$g(\theta_n^{P*}) \approx g(\hat{\theta}_n) + \frac{1}{\sqrt{n}} g'_{\theta_0}(Z).$$

Moreover, this distribution coincides with the asymptotic distribution of the plug-in estimator  $g(\hat{\theta}_n)$ , by a standard delta-method argument.

If  $g'_{\theta_0}$  is nonlinear the limiting distribution of  $\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n))$  becomes a nonlinear transformation of  $Z$ . This nonlinear transformation need not be Gaussian, and need not be centered at zero (even if  $Z$  is). Moreover, the nonlinear transformation  $g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)$  is different from the asymptotic distribution of the plug-in estimator  $g(\hat{\theta}_n)$  which is given by  $g'_{\theta_0}(Z)$ .<sup>11</sup> Thus, one can say that for directionally differentiable functions:

$$g(\theta_n^{P*}) \approx g(\hat{\theta}_n) + \frac{1}{\sqrt{n}}(g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n)), \text{ where } Z_n = \sqrt{n}(\hat{\theta}_n - \theta_0).$$

---

<sup>11</sup>This follows from an application of the delta-method for directionally differentiable functions in Shapiro (1991) or from Proposition 1 in Dümbgen (1993).

### 3. ILLUSTRATION OF MAIN RESULTS FOR $|\theta|$

The main result of this paper, Theorem 1, can be illustrated in the following simple parametric environment. Let  $X^n = (X_1, \dots, X_n)$  be an i.i.d. sample of size  $n$  from the statistical model:

$$X_i \sim \mathcal{N}(\theta, 1).$$

Consider the following family of priors for  $\theta$ :

$$\theta \sim N(0, (1/\lambda^2)),$$

where the precision parameter satisfies  $\lambda^2 > 0$ . The transformation of interest is the absolute value function:

$$g(\theta) = |\theta|.$$

It is first shown that when  $\theta_0 = 0$  this environment satisfies Assumptions 1 to 4. Then, the bootstrap and posterior distributions for  $g(\theta)$  are explicitly computed and compared.

RELATION TO MAIN ASSUMPTIONS: The transformation  $g$  is Lipschitz continuous and differentiable everywhere, except at  $\theta_0 = 0$ . At this particular point in the parameter space,  $g$  has directional derivative  $g'_0(h) = |h|$ . Thus, Assumption 1 is satisfied.

We consider the Maximum Likelihood estimator, which is given by  $\hat{\theta}_n = (1/n) \sum_{i=1}^n X_i$  and so  $\sqrt{n}(\hat{\theta}_n - \theta) \sim Z \sim \mathcal{N}(0, 1)$ . This means that Assumption 2 is satisfied.

This environment is analytically tractable so the distributions of  $\theta_n^{P*}$  and  $\theta_n^{B*}$  can be computed explicitly. The posterior distribution for  $\theta$  is given by:

$$\theta_n^{P*} | X^n \sim \mathcal{N}\left(\frac{n}{n + \lambda^2} \hat{\theta}_n, \frac{1}{n + \lambda^2}\right),$$

which implies that:

$$\sqrt{n}(\theta_n^{P*} - \hat{\theta}_n) | X^n \sim \mathcal{N}\left(\frac{\lambda^2}{n + \lambda^2} \sqrt{n} \hat{\theta}_n, \frac{n}{n + \lambda^2}\right).$$

Consequently,

$$\beta\left(\sqrt{n}(\theta_n^{P*} - \hat{\theta}_n), \mathcal{N}(0, 1); X^n\right) \xrightarrow{P} 0.$$

This implies that under,  $\theta_0 = 0$ , the first part of Assumption 3 holds.<sup>12</sup>

Second, consider a parametric bootstrap for the sample mean,  $\widehat{\theta}_n$ . We decided to focus on the parametric bootstrap to keep the exposition as simple as possible. The parametric bootstrap is implemented by generating a large number of draws  $(x_1^j, \dots, x_n^j)$ ,  $j = 1, \dots, J$  from the model

$$x_i^j \sim \mathcal{N}(\widehat{\theta}_n, 1), \quad i = 1, \dots, n,$$

recomputing the ML estimator for each of the draws. This implies that the bootstrap distribution of  $\widehat{\theta}_n$  is given by:

$$\theta_n^{B*} \sim \mathcal{N}(\widehat{\theta}_n, 1/n),$$

and so, for the parametric bootstrap it is straightforward to see that:

$$\beta\left(\sqrt{n}(\theta_n^{B*} - \widehat{\theta}_n), \mathcal{N}(0, 1); X^n\right) = 0.$$

This means that the second part of Assumption 3 holds.

Finally, Remark 2 implies that the first part of Assumption 4 is verified with:

$$h_0(Z, X^n) = g'_0(Z + Z_n) - g'_0(Z_n) = |Z + Z_n| - |Z_n|.$$

The (conditional) p.d.f. of  $Y \equiv g'_0(Z + Z_n) - g'_0(Z_n) = |Z + Z_n| - |Z_n|$  is that of a folded normal (shifted by minus  $|Z_n|$ ). Therefore:

$$F_0(y|X^n) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y + |Z_n| - Z_n)^2\right) + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y + |Z_n| + Z_n)^2\right),$$

this expression follows by direct computation or by replacing  $\cosh(x)$  in equa-

---

<sup>12</sup>The last equation follows from the fact that for two Gaussian real-valued random variables  $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$  and  $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$  we have that:

$$\left|\mathbb{E}[f(X)] - \mathbb{E}[f(Y)]\right| \leq \sqrt{\frac{2}{\pi}} \left|\sigma_1^2 - \sigma_2^2\right| + \left|\mu_1 - \mu_2\right|.$$

Therefore:

$$\beta\left(\sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n), \mathcal{N}(0, 1); X^n\right) \leq \sqrt{\frac{2}{\pi}} \left|\frac{n}{n + \lambda^2} - 1\right| + \left|\frac{\lambda^2}{n + \lambda^2} \sqrt{n\widehat{\theta}_n}\right|.$$

tion 29.41 in p. 453 in Johnson, Kotz, and Balakrishnan (1995) by  $(1/2)(\exp(x) + \exp(-x))$ . Note that:

$$F_0(y|X^n) \leq \sqrt{\frac{2}{\pi}},$$

which implies that the second part of Assumption 4 holds. To see this, take  $y_1 > y_2$ . Note that:

$$F_\theta(y_1|X^n) - F_\theta(y_2|X^n) = \int_{y_2}^{y_1} h(y|X^n)dy \leq (y_1 - y_2)\sqrt{\frac{2}{\pi}}.$$

An analogous argument for the case in which  $y_1 \leq y_2$  implies that the second part of Assumption 4 is verified.

ASYMPTOTIC BEHAVIOR OF POSTERIOR INFERENCE FOR  $g(\theta) = |\theta|$ : Since Assumptions 1 to 4 are satisfied, Theorem 1 holds.

In this example the posterior distribution of  $g(\theta_P^*)|X^n$  is given by:

$$\left| \frac{1}{\sqrt{n + \lambda^2}}Z^* + \frac{n}{n + \lambda^2}\hat{\theta}_n \right|, \quad Z^* \sim \mathcal{N}(0, 1)$$

and therefore  $\sqrt{n}(g(\theta_P^*) - g(\hat{\theta}_n))$  can be written as :

$$(3.1) \quad \left| \frac{\sqrt{n}}{\sqrt{n + \lambda^2}}Z^* + \frac{n}{n + \lambda^2}\sqrt{n}\hat{\theta}_n \right| - \left| \sqrt{n}\hat{\theta}_n \right|, \quad Z^* \sim \mathcal{N}(0, 1).$$

Theorem 1 shows that when  $\theta_0 = 0$  and  $n$  is large enough, this expression can be approximated in the Bounded Lipschitz metric in probability by:

$$(3.2) \quad \left| Z + Z_n \right| - \left| Z_n \right| = \left| Z + \sqrt{n}\hat{\theta}_n \right| - \left| \sqrt{n}\hat{\theta}_n \right|, \quad Z \sim \mathcal{N}(0, 1).$$

Observe that at  $\theta_0 = 0$  the sampling distribution of the plug-in ML estimator for  $|\theta|$  is given by:

$$\sqrt{n}(|\hat{\theta}_n| - |\theta_0|) \sim |Z|.$$

Thus, the approximate distribution of the posterior differs from the asymptotic distribution of the plug-in ML estimator and the typical Gaussian approximation for the posterior will not be appropriate.

ASYMPTOTIC BEHAVIOR OF PARAMETRIC BOOTSTRAP INFERENCE FOR  $g(\theta) = |\theta|$ :

The parametric bootstrap distribution of  $|\hat{\theta}_n|$ , centered and scaled, is simply given by:

$$|Z + \sqrt{n}\hat{\theta}_n| - |\sqrt{n}\hat{\theta}_n|, \quad Z \sim N(0, 1),$$

which implies that posterior distribution of  $|\theta|$  and the bootstrap distribution of  $|\hat{\theta}_n|$  are asymptotically equivalent.

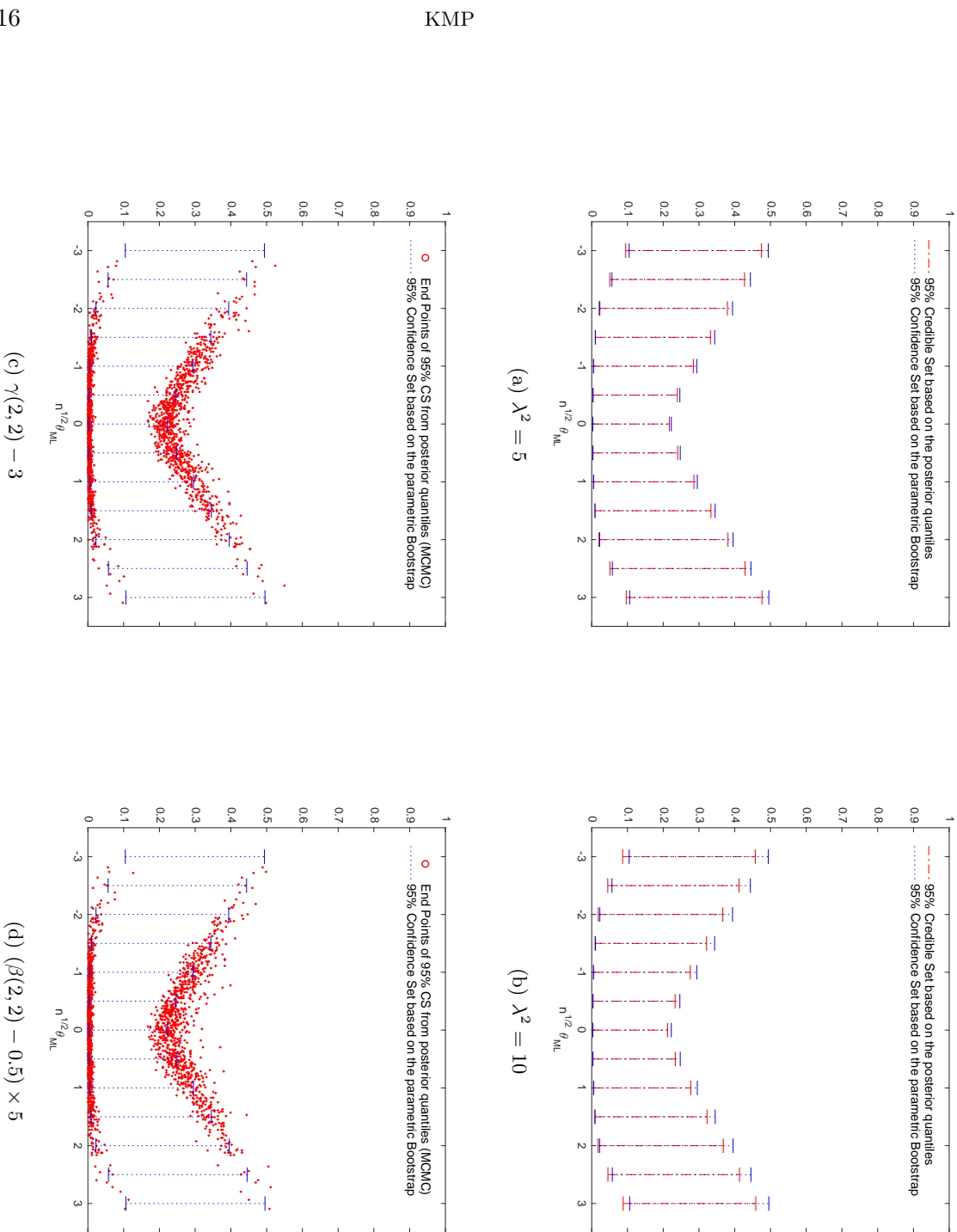
GRAPHICAL INTERPRETATION OF THEOREM 1: One way to illustrate Theorem 1 is to compute the 95% credible sets for  $|\theta|$  when  $\theta_0 = 0$  using the quantiles of the posterior. We can then compare the 95% credible sets to the 95% confidence sets from the bootstrap distribution (we have shown that Theorem 1 and Assumption 4 imply that, in large samples, these quantiles are close to each other in probability).

Observe from (3.2) that the approximation to the centered and scaled posterior and bootstrap distributions depends on the data via  $\sqrt{n}\hat{\theta}_n$ . Thus, in Figure 1 we report the 95% credible and confidence sets for data realizations  $\sqrt{n}\hat{\theta}_n \in [-3, 3]$ . In all four plots the bootstrap confidence sets are computed using the parametric bootstrap. Posterior credible sets are presented for four different priors for  $\theta$ :  $\mathcal{N}(0, 1/5)$ ,  $\mathcal{N}(0, 1/10)$ ,  $\gamma(2, 2) - 3$  and  $(\beta(2, 2) - 0.5) \times 5$ . The posterior for the first two priors is obtained using the expression in (3.1), while the posterior for the last two priors is obtained using a the Metropolis-Hastings algorithm (Geweke (2005), p. 122).

COVERAGE OF CREDIBLE SETS: In this example, the two-sided confidence set based on the quantiles of the bootstrap distribution of  $|\hat{\theta}_n|$  fails to cover  $|\theta|$  when  $\theta = 0$ . Theorem 2 showed that the two-sided credible sets based on the quantiles of the posterior should exhibit the same problem. This is illustrated in Figure 2. Thus, a frequentist cannot presume that a credible set for  $|\theta|$  based on the quantiles of the posterior will deliver a desired level of coverage.

As Liu, Gelman, and Zheng (2013) observe, although it is common to report credible sets based on the  $\alpha/2$  and  $1 - \alpha/2$  quantiles of the posterior, a Bayesian might find these credible sets unsatisfactory. In this problem, it is perhaps more natural to consider one-sided credible sets or Highest Posterior Density sets. In the online Appendix B we consider an alternative example,  $g(\theta) = \max\{\theta_1, \theta_2\}$ , where the decision between two-sided and one-sided credible sets is less obvious, but the two-sided credible set still experiences the same problem as the bootstrap.

Figure 1: 95% Credible Sets for  $|\theta|$  and 95% Parametric Bootstrap Confidence Intervals



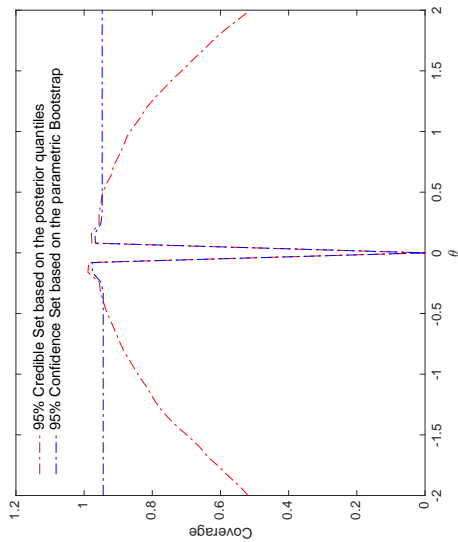
KMP

16

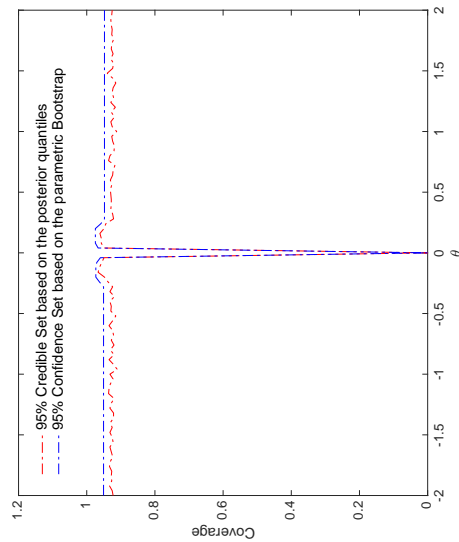
DESCRIPTION OF FIGURE 1: 95% Credible Sets for  $|\theta|$  obtained from four different priors and evaluated at different realizations of the data ( $n = 100$ ). (BLUE, DOTTED LINE) 95% confidence intervals based on the quantiles of the bootstrap distribution  $|N(\hat{\theta}_n, 1/n)|$ . The bootstrap distribution only depends on the data through  $\hat{\theta}_n$ . (RED, DOTTED LINE) 95% credible sets based on the closed-form solution for the posterior. (RED, CIRCLES) 95% credible sets based on Matlab's MCMC program (computed for a 1,000 possible data sets from a standard normal model).



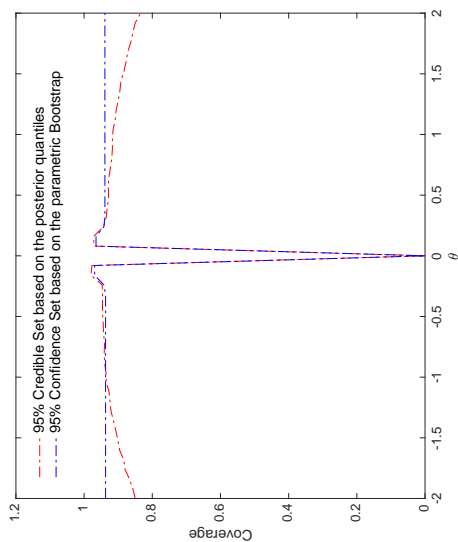
Figure 2: Coverage probability of 95% Credible Sets and Parametric Bootstrap Confidence Intervals for  $|\theta|$



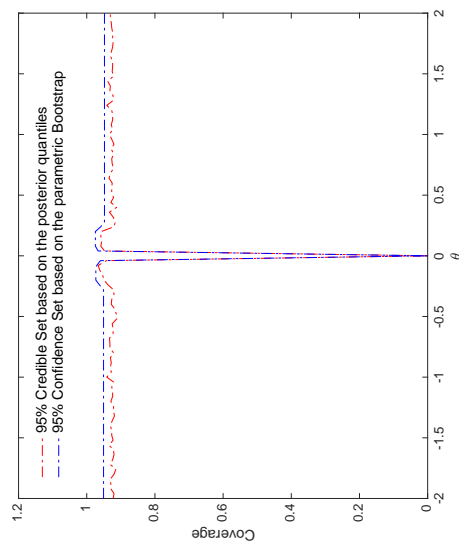
(a)  $\lambda^2 = 5$



(b)  $\lambda^2 = 10$



(c)  $\gamma(2, 2) - 3$



(d)  $\beta(2, 2) - 0.5 \times 5$

DESCRIPTION OF FIGURE 2: Coverage probability of 95% bootstrap confidence intervals and 95% Credible Sets for  $|\theta|$  obtained from four different priors and evaluated at different realizations of the data ( $n = 100$ ). (BLUE, DOTTED LINE) Coverage probability of 95% confidence intervals based on the quantiles of the bootstrap distribution  $|\mathcal{N}(\hat{\theta}_n, 1/n)|$ . (RED, DOTTED LINE) 95% credible sets based on quantiles of the posterior. Cases (a) and (b) use the closed form expression for the posterior. Cases (c) and (d) use Matlab's MCMC program.

## 4. CONCLUSION

This paper studied the asymptotic behavior of the posterior distribution of parameters of the form  $g(\theta)$ , where  $g(\cdot)$  is Lipschitz continuous but possibly nondifferentiable. We have shown that the bootstrap distribution of  $g(\hat{\theta}_n)$  and the posterior of  $g(\theta)$  are asymptotically equivalent.

One implication from our results is that Bayesians can interpret bootstrap inference for  $g(\theta)$  as approximately valid posterior inference in large samples. In fact, Bayesians can use bootstrap draws to conduct approximate posterior inference for  $g(\theta)$  whenever bootstrapping  $g(\hat{\theta}_n)$  is more convenient than MCMC sampling. This reinforces observations in the statistics literature noting that by “perturbing the data, the bootstrap approximates the Bayesian effect of perturbing the parameters” (Hastie, Tibshirani, Friedman, and Franklin (2005), p. 236).<sup>13</sup>

Another implication from our main result—combined with known results about bootstrap inconsistency—is that it takes only mild departures from differentiability (such as directional differentiability) to make the posterior distribution of  $g(\theta)$  behave differently than the limit of  $\sqrt{n}(g(\hat{\theta}_n) - g(\theta))$ . We showed that whenever nondifferentiability causes a bootstrap confidence set to cover  $g(\theta)$  less often than desired, a credible set based on the quantiles of the posterior will have distorted frequentist coverage as well.

---

<sup>13</sup>Our results also provide a better understanding of what type of statistics could preserve, in large samples, the equivalence between bootstrap and posterior resampling methods, a question that have been explored by Lo (1987).

## REFERENCES

- ANDREWS, D. W. (2000): “Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space,” *Econometrica*, 68, 399–405.
- BALKE, A. AND J. PEARL (1997): “Bounds on treatment effects from studies with imperfect compliance,” *Journal of the American Statistical Association*, 92, 1171–1176.
- BERAN, R. (1997): “Diagnosing bootstrap success,” *Annals of the Institute of Statistical Mathematics*, 49, 1–24.
- CASTILLO, I. AND J. ROUSSEAU (2015): “A Bernstein–von Mises theorem for smooth functionals in semiparametric models,” *The Annals of Statistics*, 43, 2353–2383.
- DASGUPTA, A. (2008): *Asymptotic theory of statistics and probability*, Springer Verlag.
- DUDLEY, R. (2002): *Real Analysis and Probability*, vol. 74, Cambridge University Press.
- DÜMBGEN, L. (1993): “On nondifferentiable functions and the bootstrap,” *Probability Theory and Related Fields*, 95, 125–140.
- EATON, M. L. AND D. E. TYLER (1991): “On Wielandt’s inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix,” *The Annals of Statistics*, 260–271.
- EFRON, B. (2012): “Bayesian inference and the parametric bootstrap,” *The Annals of Applied Statistics*, 6, 1971.
- EVANS, L. C. AND R. F. GARIEPY (2015): *Measure theory and fine properties of functions*, CRC press.
- FANG, Z. AND A. SANTOS (2015): “Inference on Directionally Differentiable Functions,” *Working paper, University of California at San Diego*.
- FAUST, J. (1998): “The Robustness of Identified VAR Conclusions about Money,” in *Carnegie-Rochester Conference Series on Public Policy*, Elsevier, vol. 49, 207–244.
- FIELLER, E. C. (1954): “Some Problems in Interval Estimation,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 175–185.
- GAFAROV, B., M. MEIER, AND J. L. MONTIEL OLEA (2015): “Delta-Method inference for a class of set-identified SVARs,” *Working paper, New York University*.
- GEWEKE, J. (2005): *Contemporary Bayesian econometrics and statistics*, vol. 537, John Wiley & Sons.
- GHOSAL, S., J. K. GHOSH, AND T. SAMANTA (1995): “On convergence of posterior distributions,” *The Annals of Statistics*, 2145–2152.

- GIACOMINI, R. AND T. KITAGAWA (2015): “Robust Inference about partially identified SVARs,” *Working Paper, University College London*.
- HAILE, P. A. AND E. TAMER (2003): “Inference with an incomplete model of English auctions,” *Journal of Political Economy*, 111, 1–51.
- HANSEN, B. E. (2015): “Regression kink with an unknown threshold,” *Journal of Business & Economic Statistics*.
- HASTIE, T., R. TIBSHIRANI, J. FRIEDMAN, AND J. FRANKLIN (2005): “The elements of statistical learning: data mining, inference and prediction,” *The Mathematical Intelligencer*, 27, 83–85.
- HIRANO, K. AND J. R. PORTER (2012): “Impossibility results for nondifferentiable functionals,” *Econometrica*, 80, 1769–1790.
- HONG, H. AND J. LI (2015): “The numerical delta-method,” *Working Paper, Stanford University*.
- HOROWITZ, J. L. (2001): “The bootstrap,” *Handbook of econometrics*, 5, 3159–3228.
- JOHNSON, N., S. KOTZ, AND N. BALAKRISHNAN (1995): “Continuous Univariate Distributions, Vol. 2. 1995,” .
- LIU, Y., A. GELMAN, AND T. ZHENG (2013): “Simulation-efficient shortest probability intervals,” *arXiv preprint arXiv:1302.2142*.
- LO, A. Y. (1987): “A large sample study of the Bayesian bootstrap,” *The Annals of Statistics*, 360–375.
- MANSKI, C. F. (1990): “Nonparametric bounds on treatment effects,” *The American Economic Review*, 80, 319–323.
- (2004): “Statistical treatment rules for heterogeneous populations,” *Econometrica*, 72, 1221–1246.
- NEWKEY, W. AND D. MCFADDEN (1994): “Large sample estimation and hypothesis testing,” *Handbook of econometrics*, 2111–2245.
- RUBIN, D. B. (1981): “The Bayesian Bootstrap,” *The Annals of Statistics*, 9, 130–134.
- SHAPIRO, A. (1990): “On concepts of directional differentiability,” *Journal of optimization theory and applications*, 66, 477–487.
- (1991): “Asymptotic analysis of stochastic programs,” *Annals of Operations Research*, 30, 169–186.
- VAN DER VAART, A. (2000): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
- VAN DER VAART, A. AND J. WELLNER (1996): *Weak Convergence and Empirical Processes*,

Springer, New York.

## APPENDIX A: MAIN THEORETICAL RESULTS.

## A.1. Proof of Theorem 1

**LEMMA 1** *Suppose that Assumption 1 holds. Suppose that  $\theta_n^*$  is a random variable satisfying:*

$$\sup_{f \in BL(1)} \left| \mathbb{E}[f(Z_n^*) | X^n] - \mathbb{E}[f(Z^*)] \right| \xrightarrow{P} 0,$$

where  $Z_n^* = \sqrt{n}(\theta_n^* - \hat{\theta}_n)$  and  $Z^*$  is a random variable independent of  $X^n = (X_1, \dots, X_n)$  for every  $n$ . Then,

$$(A.1) \quad \sup_{f \in BL(1)} \left| \mathbb{E}[f(\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n))) | X^n] \right. \\ \left. - \mathbb{E}[f(\sqrt{n}(g(\theta_0 + Z^*/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))) | X^n] \right| \xrightarrow{P} 0,$$

where  $\theta_0$  is the parameter that generated the data and  $Z_n = \sqrt{n}(\hat{\theta}_n - \theta_0)$ .

**PROOF:** By Assumption 1,  $g$  is Lipschitz continuous. Define  $\Delta_n(a) = \sqrt{n}(g(\theta_0 + a/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))$ . Observe that  $\Delta_n(\cdot)$  is Lipschitz since:

$$|\Delta_n(a) - \Delta_n(b)| = |\sqrt{n}(g(\theta_0 + a/\sqrt{n} + Z_n/\sqrt{n}) - g(\theta_0 + b/\sqrt{n} + Z_n/\sqrt{n}))| \\ \leq c\|a - b\|, \\ \text{(by Assumption 1).}$$

Define  $\tilde{c} = \max\{c, 1\}$ . Then, the function  $(f \circ \Delta_n)/\tilde{c}$  is an element of  $BL(1)$  (if  $f$  is). Consequently,

$$\left| \mathbb{E}[f(\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n))) | X^n] \right. \\ \left. - \mathbb{E}[f(\sqrt{n}(g(\theta_0 + Z^*/\sqrt{n} + Z_n/\sqrt{n}) - g(\hat{\theta}_n))) | X^n] \right| \\ = \tilde{c} \left| \mathbb{E} \left[ \frac{f \circ \Delta_n(Z_n^*)}{\tilde{c}} \mid X^n \right] - \mathbb{E} \left[ \frac{f \circ \Delta_n(Z^*)}{\tilde{c}} \mid X^n \right] \right|, \\ \text{(since } \theta_n^* = \theta_0 + Z_n^*/\sqrt{n} + Z_n/\sqrt{n}) \\ \leq \tilde{c} \sup_{f \in BL(1)} \left| \mathbb{E}[f(Z_n^*) | X^n] - \mathbb{E}[f(Z^*) | X^n] \right|, \\ \text{(since } (f \circ \Delta_n)/\tilde{c} \in BL(1)).$$

*Q.E.D.*

**PROOF OF THEOREM 1:** Theorem 1 follows from Lemma 1. Note first that Assumptions 1, 2 and 3 imply that the assumptions of Lemma 1 are verified for both  $\theta_n^{P*}$  and  $\theta_n^{B*}$ . Note then that:

$$\sup_{f \in BL(1)} \left| \mathbb{E}[f(\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n))) | X^n] - \mathbb{E}[f(\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n))) | X^n] \right| \\ \leq \sup_{f \in BL(1)} \left| \mathbb{E}[f(\sqrt{n}(g(\theta_n^{P*}) - g(\hat{\theta}_n))) | X^n] \right|$$

$$\begin{aligned}
& - \mathbb{E} \left[ f \left( \sqrt{n} \left( g \left( \theta_0 + \frac{Z}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}} \right) - g(\hat{\theta}_n) \right) \right) \mid X^n \right] \\
& + \sup_{f \in BL(1)} \left| \mathbb{E} [f(\sqrt{n}(g(\theta_n^{B*}) - g(\hat{\theta}_n))) \mid X^n] \right. \\
& \quad \left. - \mathbb{E} \left[ f \left( \sqrt{n} \left( g \left( \theta_0 + \frac{Z}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}} \right) - g(\hat{\theta}_n) \right) \right) \mid X^n \right] \right|.
\end{aligned}$$

Lemma 1 implies that both terms converge to zero in probability.

*Q.E.D.*

## A.2. Proof of Theorem 2

We start by establishing a Lemma based on a high-level assumption implied by the second part of Assumption 4. In what follows we use  $\mathbb{P}^Z$  to denote the distribution of the random variable  $Z$  (which is independent of the data  $X^n$  for every  $n$ ).

**ASSUMPTION 5** The function  $h_\theta(Z, X^n)$  is such that for all positive  $(\epsilon, \delta)$  there exists  $\zeta(\epsilon, \delta) > 0$  and  $N(\epsilon, \delta)$  for which:

$$\mathbb{P}_\theta \left( \sup_{c \in \mathbb{R}} \mathbb{P}^Z \left( c - \zeta(\epsilon, \delta) \leq h_\theta(Z, X^n) \leq c + \zeta(\epsilon, \delta) \mid X^n \right) > \epsilon \right) < \delta.$$

provided  $n \geq N(\epsilon, \delta)$ .

To see that Assumption 5 is implied by the second part of Assumption 4 simply note that:

$$\mathbb{P}^Z \left( c - \zeta(\epsilon, \delta) \leq h_\theta(Z, X^n) \leq c + \zeta(\epsilon, \delta) \mid X^n \right),$$

equals:

$$F_\theta(c + \zeta(\epsilon, \delta) | X^n) - F_\theta(c - \zeta(\epsilon, \delta) | X^n) \leq 2\zeta(\epsilon, \delta)k.$$

Note that the last inequality holds since, by assumption,  $F_\theta(y | X^n)$  is Lipschitz continuous—for almost every  $X_n$  for every  $n$ —with a constant  $k$  that does not depend on  $X^n$ . By choosing  $\zeta(\epsilon, \delta)$  equal to  $\epsilon/4k$ , then

$$\mathbb{P}^Z \left( c - \zeta(\epsilon, \delta) \leq h_\theta(Z, X^n) \leq c + \zeta(\epsilon, \delta) \mid X^n \right) \leq \frac{\epsilon}{2},$$

for every  $c$ , implying that Assumption 5 holds.

We now show that any random variable satisfying the weak convergence assumption in the first part of Assumption 4 has a conditional  $\alpha$ -quantile that—with high probability—lies in between the conditional  $(\alpha - \epsilon)$  and  $(\alpha + \epsilon)$ -quantiles of the limiting distribution.

**LEMMA 2** Let  $\theta_n^*$  denote a random variable whose distribution,  $P^*$ , depends on  $X^n = (X_1, \dots, X_n)$  and let  $Z$  be the limiting distribution of  $Z_n \equiv \sqrt{n}(\hat{\theta}_n - \theta)$  as defined in Assumption 2. Suppose that

$$\beta(\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n)), h_\theta(Z, X^n); X^n) \xrightarrow{P} 0.$$

Define  $c_\alpha^*(X^n)$  as the critical value such that:

$$c_\alpha^*(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n)) \leq c \mid X^n) \geq \alpha\}.$$

Suppose  $h_\theta(Z, X^n)$  satisfies Assumption 5 and define  $c_\alpha(X^n)$  as:

$$\mathbb{P}^Z (h_\theta(Z, X^n) \leq c_\alpha(X^n) \mid X^n) = \alpha.$$



Then for any  $0 < \epsilon < \alpha$  and  $\delta > 0$  there exists  $N(\epsilon, \delta)$  such that for  $n > N(\epsilon, \delta)$ :

$$\mathbb{P}_\theta(c_{\alpha-\epsilon}(X^n) \leq c_\alpha^*(X^n) \leq c_{\alpha+\epsilon}(X^n)) \geq 1 - \delta.$$

PROOF: We start by deriving a convenient bound for the difference between the conditional distributions of  $\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n))$  and the distribution of  $h_\theta(Z, X^n)$ . Define the random variables:

$$W_n^* \equiv \sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n)), \quad Y_n^* \equiv h_\theta(Z, X^n).$$

Denote by  $P_W^n$  and  $P_Y^n$  the probabilities that each of these random variables induce over the real line. Let  $c \in \mathbb{R}$  be some constant. By applying Lemma 5 in Appendix A.4 to the set  $A = (-\infty, c)$  it follows that for any  $\zeta > 0$ :

$$\begin{aligned} & |P_W^n((-\infty, c)|X^n) - P_Y^n((-\infty, c)|X^n)| \\ & \leq \frac{1}{\zeta} \beta(W_n^*, Y_n^*; X^n) + \min\{(P_Y^n(A^\zeta \setminus A|X^n), P_Y^n((A^c)^\zeta \setminus A^c|X^n))\} \\ & = \frac{1}{\zeta} \beta(W_n^*, Y_n^*; X^n) + \min\{P_Y^n([c, c + \zeta]|X^n), P_Y^n([c - \zeta, c]|X^n)\} \\ & \leq \frac{1}{\zeta} \beta(W_n^*, Y_n^*; X^n) + \mathbb{P}^Z(c - \zeta \leq h_\theta(Z, X^n) \leq c + \zeta | X^n), \end{aligned}$$

where for any set  $A$ , we define  $A^\delta \equiv \{y \in \mathbb{R}^k : \|x - y\| < \delta \text{ for some } x \in A\}$  (see Lemma 5). Therefore: That is:

$$\begin{aligned} & |\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n)) \leq c | X^n) - \mathbb{P}^Z(h_\theta(Z, X^n) \leq c | X^n)| \\ & \leq \frac{1}{\zeta} \beta(\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n)), h_\theta(Z, X^n); X^n) \\ & \quad + \sup_{c \in \mathbb{R}} \mathbb{P}^Z(c - \zeta \leq h_\theta(Z, X^n) \leq c + \zeta | X^n) \end{aligned}$$

We use this relation between the conditional c.d.f. of  $\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n))$  and the conditional c.d.f. of  $h_\theta(Z, X^n)$  to show that quantiles of these distributions should be close to each other.

To simplify the notation, define the functions:

$$\begin{aligned} A_1(\zeta, X^n) & \equiv \frac{1}{\zeta} \beta(\sqrt{n}(g(\theta_n^*) - g(\hat{\theta}_n)), h_\theta(Z, X^n); X^n), \\ A_2(\zeta, X^n) & \equiv \sup_{c \in \mathbb{R}} \mathbb{P}^Z(c - \zeta \leq h_\theta(Z, X^n) \leq c + \zeta | X^n). \end{aligned}$$

Observe that if the data  $X^n$  were such that  $A_1(\zeta, X^n) \leq \epsilon/2$  and  $A_2(\zeta, X^n) \leq \epsilon/2$  then for any

$c \in \mathbb{R}$ :

$$\begin{aligned} & |\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c | X^n) - \mathbb{P}^Z(h_\theta(Z, X^n) \leq c | X^n)| \\ & \leq A_1(\zeta, X^n) + A_2(\zeta, X^n) \\ & < \epsilon. \end{aligned}$$

This would imply that for any  $c \in \mathbb{R}$ :

$$(A.2) \quad -\epsilon < \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c | X^n) - \mathbb{P}^Z(h_\theta(Z, X^n) \leq c | X^n) < \epsilon.$$

We now show that this inequality implies that:

$$c_{\alpha-\epsilon}(X^n) \leq c_\alpha^*(X^n) \leq c_{\alpha+\epsilon}(X^n),$$

whenever  $X^n$  is such that  $A_1(\zeta, X^n) \leq \epsilon/2$  and  $A_2(\zeta, X^n) \leq \epsilon/2$ . To see this, evaluate equation (A.2) at  $c_{\alpha+\epsilon}(X^n)$ . This implies that:

$$-\epsilon < \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c_{\alpha+\epsilon}(X^n) | X^n) - (\alpha + \epsilon).$$

Consequently:

$$c_{\alpha+\epsilon}(X^n) \in \{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c | X^n) \geq \alpha\}.$$

Since:

$$c_\alpha^*(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c | X^n) \geq \alpha\},$$

it follows that:

$$c_\alpha^*(X^n) \leq c_{\alpha+\epsilon}(X^n).$$

To obtain the other inequality, evaluate equation (A.2) at  $c_{\alpha-\epsilon}(X^n)$ . This implies that:

$$\mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c_{\alpha-\epsilon}(X^n) | X^n) - (\alpha - \epsilon) < \epsilon.$$

Note that  $c_{\alpha-\epsilon}(X^n)$  is a lower bound of the set:

$$(A.3) \quad \{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c | X^n) \geq \alpha\}.$$

If this were not the case, there would exist  $c^*$  in the set above such that  $c^* < c_{\alpha-\epsilon}(X^n)$ . As a consequence, the monotonicity of the c.d.f would then imply that:

$$\alpha \leq \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c^* | X^n) \leq \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c_{\alpha-\epsilon}(X^n) | X^n) < \alpha,$$

which would imply that  $\alpha < \alpha$ ; a contradiction. Therefore,  $c_{\alpha-\epsilon}(X^n)$  is indeed a lower bound for the set in (A.3) and, consequently:

$$c_{\alpha-\epsilon}(X^n) \leq c_\alpha^*(X^n) \equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^*(\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n)) \leq c | X^n) \geq \alpha\}.$$

This shows that whenever the data  $X^n$  is such that  $A_1(\zeta, X^n) \leq \epsilon/2$  and  $A_2(\zeta, X^n) \leq \epsilon/2$

$$c_{\alpha-\epsilon}(X^n) \leq c_\alpha^*(X^n) \leq c_{\alpha+\epsilon}(X^n).$$

To finish the proof, note that by Assumption 5 there exists  $\zeta^* \equiv \zeta(\epsilon/2, \delta/2)$  and  $N(\epsilon/2, \delta/2)$  that guarantees that if  $n > N(\epsilon/2, \delta/2)$ :

$$\mathbb{P}_\theta^n(A_2(\zeta^*, X^n) > \epsilon/2) < \delta/2.$$

Also, by the convergence assumption of this Lemma, there is  $N(\zeta^*, \epsilon/2, \delta/2)$  such that for  $n > N(\zeta^*, \epsilon/2, \delta/2)$ :

$$\mathbb{P}_\theta^n(A_1(\zeta^*, X^n) > \epsilon/2) < \delta/2.$$

It follows that for  $n > \max\{N(\zeta^*, \epsilon/2, \delta/2), N(\epsilon/2, \delta/2)\} \equiv N(\epsilon, \delta)$

$$\begin{aligned} \mathbb{P}_\theta(c_{\alpha-\epsilon}(X^n) \leq c_\alpha^*(X^n) \leq c_{\alpha+\epsilon}(X^n)) & \\ & \geq \mathbb{P}_\theta(A_1(\zeta^*, X^n) < \epsilon/2 \text{ and } A_2(\zeta^*, X^n) < \epsilon/2) \\ & = 1 - \mathbb{P}_\theta(A_1(\zeta^*, X^n) > \epsilon/2 \text{ or } A_2(\zeta^*, X^n) > \epsilon/2) \\ & \geq 1 - \mathbb{P}_\theta(A_1(\zeta^*, X^n) > \epsilon/2) - \mathbb{P}_\theta(A_2(\zeta^*, X^n) > \epsilon/2) \\ & \geq 1 - \delta \end{aligned}$$

*Q.E.D.*

We have shown that if  $\sqrt{n}(g(\theta_n^*) - g(\widehat{\theta}_n))$  is any random variable satisfying the assumptions of Lemma 2, its conditional  $\alpha$ -quantile lies—with high probability—between the conditional  $(\alpha - \epsilon)$  and  $(\alpha + \epsilon)$  quantiles of the limiting distribution  $h_\theta(Z, X^n)$ . The next Lemma considers the case in which  $\theta_n^*$  is either  $\theta_n^{B^*}$  or  $\theta_n^{P^*}$  and characterizes the asymptotic behavior of the c.d.f. of  $\sqrt{n}(g(\widehat{\theta}_n) - g(\theta))$  evaluated at bootstrap and posterior quantiles. The main result is that the c.d.f. evaluated at the  $\alpha$ -bootstrap quantile is—in large samples—close to same c.d.f. evaluated at the  $(\alpha - \epsilon)$  and  $(\alpha + \epsilon)$  posterior quantiles. We note that this result could not be obtained directly from the fact that the bootstrap and posterior quantiles converge in probability to each other, as some additional regularity in the limiting distribution is needed. This is why it was important to establish Lemma 2 before the following Lemma.

**LEMMA 3** *Suppose that Assumptions 1-4 hold. Fix  $\alpha \in (0, 1)$ . Let  $c_\alpha^B(X^n)$  and  $c_\alpha^P(X^n)$  denote critical values satisfying:*

$$\begin{aligned} c_\alpha^{B^*}(X^n) &\equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^{B^*}(\sqrt{n}(g(\theta_n^{B^*}) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \alpha\}, \\ c_\alpha^{P^*}(X^n) &\equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^{P^*}(\sqrt{n}(g(\theta_n^{P^*}) - g(\widehat{\theta}_n)) \leq c \mid X^n) \geq \alpha\}. \end{aligned}$$

Then, for any  $0 < \epsilon < \alpha$  and  $\delta > 0$  there exists  $N(\epsilon, \delta)$  such that for all  $n > N(\epsilon, \delta)$ :

$$(A.4) \quad \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B^*}(X^n)) \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P^*}(X^n)) + \delta,$$

$$(A.5) \quad \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{P^*}(X^n)) \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{B^*}(X^n)) - \delta.$$

PROOF: Let  $\theta^*$  denote either  $\theta_n^{P^*}$  or  $\theta_n^{B^*}$ . Let  $c_\alpha(X^n)$  and  $c_\alpha^*(X^n)$  be defined as in Lemma 2. Under Assumptions 1 to 4, the conditions for Lemma 2 are satisfied. It follows that for any  $0 < \epsilon < \alpha$  and  $\delta > 0$  there exists  $N(\epsilon, \delta)$  such that for all  $n > N(\epsilon, \delta)$ :

$$\mathbb{P}_\theta(c_{\alpha+\epsilon/2}(X^n) < c_\alpha^*(X^n)) \leq \delta/2 \quad \text{and} \quad \mathbb{P}_\theta(c_\alpha^*(X^n) < c_{\alpha-\epsilon/2}(X^n)) \leq \delta/2.$$

Therefore:

$$\begin{aligned} (A.6) \quad &\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n)) \\ &= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n) \text{ and } c_{\alpha+\epsilon/2}(X^n) \geq c_\alpha^*(X^n)) \\ &\quad + \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(X^n) \text{ and } c_{\alpha+\epsilon/2}(X^n) < c_\alpha^*(X^n)) \\ &\text{(by the additivity of probability measures)} \\ &\leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \mathbb{P}_\theta(c_{\alpha+\epsilon/2}(X^n) < c_\alpha^*(X^n)) \\ &\text{(by the monotonicity of probability measures)} \\ &\leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \delta/2. \end{aligned}$$

Also, we have that:

$$\begin{aligned} (A.7) \quad &\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(X^n)) \\ &\geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(X^n) \text{ and } c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \end{aligned}$$

$$\begin{aligned}
&\geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n) \text{ and } c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \\
&= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) + \mathbb{P}_\theta(c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \\
&\quad - \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n) \text{ or } c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \\
&\quad \text{(using } P(A \cap B) = P(A) + P(B) - P(A \cup B)) \\
&\geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) - (1 - \mathbb{P}_\theta(c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(X^n))) \\
&\quad \text{(since } \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n) \text{ or } c_\alpha^*(X^n) \geq c_{\alpha-\epsilon/2}(X^n)) \leq 1) \\
&= \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) - \mathbb{P}_\theta(c_\alpha^*(X^n) < c_{\alpha-\epsilon/2}(X^n)) \\
&\geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^*(X^n)) - \delta/2.
\end{aligned}$$

Replacing  $c_\alpha^*$  by  $c_\alpha^{B^*}$  in (A.7) and  $c_\alpha^*$  by  $c_\alpha^{P^*}$  and  $\alpha$  by  $\alpha - \epsilon$  in (A.6) implies that for  $n > N(\epsilon, \delta)$ :

$$\begin{aligned}
\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(X^n)) &\geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}^{B^*}(X^n)) - \delta/2 \\
\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}(X^n)) &\leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon/2}^{P^*}(X^n)) + \delta/2.
\end{aligned}$$

Combining the previous two equations gives that for  $n > N(\epsilon, \delta)$ :

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B^*}(X^n)) \leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha-\epsilon}^{P^*}(X^n)) + \delta.$$

This establishes equation (A.4). Replacing  $\theta_n^*$  by  $\theta_n^{B^*}$  in (A.6) and replacing  $\theta_n^*$  by  $\theta_n^{P^*}$ ,  $\alpha$  by  $\alpha + \epsilon$  (A.7) implies that for  $n > N(\epsilon, \delta)$ :

$$\begin{aligned}
\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(Z_n)) &\leq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}^{B^*}(X^n)) + \delta/2 \\
\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}(Z_n)) &\geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon/2}^{P^*}(X^n)) - \delta/2
\end{aligned}$$

and combining the previous two equations gives that for  $n > N(\epsilon, \delta)$ :

$$\mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_\alpha^{B^*}(X^n)) \geq \mathbb{P}_\theta(\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \leq -c_{\alpha+\epsilon}^{P^*}(X^n)) - \delta,$$

which establishes equation (A.5).

*Q.E.D.*

PROOF OF THEOREM 2: Define, for any  $0 < \beta < 1$ , the critical values  $c_\beta^B(X^n)$  and  $c_\beta^P(X^n)$  as:

$$\begin{aligned} c_\beta^{B^*}(X^n) &\equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^{B^*}(\sqrt{n}(g(\theta_n^{B^*}) - g(\hat{\theta}_n)) \leq c \mid X^n) \geq \beta\}, \\ c_\beta^{P^*}(X^n) &\equiv \inf_c \{c \in \mathbb{R} \mid \mathbb{P}^{P^*}(\sqrt{n}(g(\theta_n^{P^*}) - g(\hat{\theta}_n)) \leq c \mid X^n) \geq \beta\}. \end{aligned}$$

Note that the critical values  $c_\beta^{B^*}(X^n)$ ,  $c_\beta^{P^*}(X^n)$  and the quantiles for  $g(\theta_n^{B^*})$  and  $g(\theta_n^{P^*})$  are related through the equation:

$$\begin{aligned} q_\beta^B(X^n) &= g(\hat{\theta}_n) + c_\beta^{B^*}(X^n)/\sqrt{n} \\ q_\beta^P(X^n) &= g(\hat{\theta}_n) + c_\beta^{P^*}(X^n)/\sqrt{n}. \end{aligned}$$

This implies that:

$$\begin{aligned} CS_n^B(1 - \alpha) &= \left[ g(\hat{\theta}_n) + c_{\alpha/2}^{B^*}(X^n)/\sqrt{n}, g(\hat{\theta}_n) + c_{1-\alpha/2}^{B^*}(X^n) \right] \\ CS_n^P(1 - \alpha - \epsilon) &= \left[ g(\hat{\theta}_n) + c_{\alpha/2+\epsilon/2}^{P^*}(X^n)/\sqrt{n}, g(\hat{\theta}_n) + c_{1-\alpha/2-\epsilon/2}^{P^*}(X^n) \right]. \end{aligned}$$

Under Assumptions 1 to 4 we can apply the previous lemma. This implies that for  $n > N(\epsilon, \delta)$

$$\begin{aligned} \mathbb{P}_\theta(g(\theta) \in CS_n^B(1 - \alpha)) &= \mathbb{P}_\theta \left( g(\theta) \in \left[ g(\hat{\theta}_n) + c_{\alpha/2}^{B^*}(X^n)/\sqrt{n}, g(\hat{\theta}_n) + c_{1-\alpha/2}^{B^*}(X^n) \right] \right) \\ &= \mathbb{P}_\theta(\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \leq -c_{\alpha/2}^{B^*}(X^n)) \\ &\quad - \mathbb{P}_\theta(\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \leq -c_{1-\alpha/2}^{B^*}(X^n)) \\ &\geq \mathbb{P}_\theta(\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \leq -c_{\alpha/2+\epsilon/2}^{P^*}(X^n)) \\ &\quad - \mathbb{P}_\theta(\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \leq -c_{1-\alpha/2-\epsilon/2}^{P^*}(X^n)) - \delta \\ &\quad \text{(Replacing } \alpha \text{ by } \alpha/2, \epsilon \text{ by } \epsilon/2 \text{ and } \delta \text{ by } \delta/2 \text{ in (A.5) and} \\ &\quad \text{replacing } \alpha \text{ by } 1 - \alpha/2, \epsilon \text{ by } \epsilon/2 \text{ and } \delta \text{ by } \delta/2 \text{ in (A.4))} \\ &= \mathbb{P}_\theta(g(\theta) \in CS_n^P(1 - \alpha - \epsilon)) - \delta \end{aligned}$$

This implies that for every  $\epsilon > 0$ :

$$1 - \alpha - d_\alpha \geq \limsup_{n \rightarrow \infty} \mathbb{P}_\theta(g(\theta) \in CS_n^B) \geq \limsup_{n \rightarrow \infty} \mathbb{P}_\theta(g(\theta) \in CS_n^P(1 - \alpha - \epsilon)),$$

which implies that

$$1 - \alpha - \epsilon - (d_\alpha - \epsilon) \geq \limsup_{n \rightarrow \infty} \mathbb{P}_\theta(g(\theta) \in CS_n^P(1 - \alpha - \epsilon)).$$

This implies that if the bootstrap fails at  $\theta$  by at least  $d_\alpha\%$  given the nominal confidence level  $(1 - \alpha)\%$ , then the confidence set based on the quantiles of the posterior will fail at  $\theta$ —by at least  $(d_\alpha - \epsilon)\%$ —given the nominal confidence level  $(1 - \alpha - \epsilon)\%$ .

A.3. Posterior Distribution of  $g(\theta^{P^*})$  under directional differentiability

**LEMMA 4** *let  $Z$  be the limiting distribution of  $Z_n \equiv \sqrt{n}(\hat{\theta}_n - \theta)$  as defined in Assumption 2. Let  $Z^*$  be a random variable independent of both  $X^n = (X_1, \dots, X_n)$  and  $Z$  and let  $\theta_0$  denote the parameter that generated the data. Suppose that  $g$  is directionally differentiable in the sense defined in Remark 2 of the main text. Then, Assumption 4 (i) holds with  $h_{\theta_0}(Z, Z_n) = g'_{\theta_0}(Z^* + Z_n) - g'_{\theta_0}(Z_n)$ .*

PROOF: We start by analyzing the limiting distribution of both:

$$\sqrt{n}(g(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}) - g(\theta_0))$$

and

$$\sqrt{n}(g(\theta_0 + \frac{Z_n}{\sqrt{n}}) - g(\theta_0))$$

as a function of  $(Z^*, Z_n)$ . Note that the delta-method for directionally differentiable functions (e.g., Theorem 2.1 in Fang and Santos (2015)) and the continuity of the directional derivative implies that jointly:

$$\begin{aligned} \sqrt{n}(g(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}) - g(\theta_0)) &\xrightarrow{d} g'_{\theta_0}(Z^* + Z) \\ g'_{\theta_0}(Z^* + Z_n) &\xrightarrow{d} g'_{\theta_0}(Z^* + Z) \\ \sqrt{n}(g(\theta_0 + Z_n/\sqrt{n}) - g(\theta_0)) &\xrightarrow{d} g'_{\theta_0}(Z) \\ g'_{\theta_0}(Z_n) &\xrightarrow{d} g'_{\theta_0}(Z) \end{aligned}$$

where  $Z$  is independent of  $Z^*$ . Note that the joint (and unconditional) convergence in distribution above implies that:

$$A_n \equiv \sqrt{n}(g(\theta_0 + \frac{Z^*}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}) - g(\hat{\theta}_n))$$

and

$$B_n \equiv g'_{\theta_0}(Z^* + Z_n) - g'_{\theta_0}(Z_n)$$

are such that  $|A_n - B_n| = o_p(1)$ , where the  $o_p(1)$  terms refers to convergence in probability unconditional on the data as a function of  $Z^*$  and  $Z_n$ .

Note that for any two random variables  $A_n$  and  $B_n$  we have that for any  $\epsilon$

$$\sup_{BL(1)} \left| \mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n] \right|$$

is bounded above by:

$$\epsilon + 2\mathbb{P}^{Z^*} [ |A_n - B_n| > \epsilon |X^n ],$$

where the probability is taken over the distribution of  $Z^*$ , denoted  $\mathbb{P}^{Z^*}$ .<sup>14</sup> Note that the unconditional convergence in probability result for  $|A_n - B_n|$  implies that:

---

<sup>14</sup>This is a common bound used in bootstrap analysis; see for example, Theorem 23.9 p. 333 in Van der Vaart (2000).

$$\mathbb{E}_\theta[\mathbb{P}^{Z^*} [|A_n - B_n| > \epsilon | X^n]] \rightarrow 0,$$

as the expectation is taken over different data realizations. Note that in light of the inequalities above we have that:

$$(A.8) \quad \mathbb{P}_\theta \left( \sup_{BL(1)} \left| \mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n] \right| > 2\epsilon \right)$$

is bounded above by:

$$\mathbb{P}_\theta \left( \epsilon + 2\mathbb{P}^{Z^*} [|A_n - B_n| > \epsilon | X^n] > 2\epsilon \right),$$

which equals

$$\mathbb{P}_\theta \left( \mathbb{P}^{Z^*} [|A_n - B_n| > \epsilon | X^n] > \epsilon/2 \right).$$

Thus, by Markov's inequality:

$$\mathbb{P}_\theta \left( \sup_{BL(1)} \left| \mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n] \right| > 2\epsilon \right) \leq 2\mathbb{E}_\theta[\mathbb{P}^{Z^*} [|A_n - B_n| > \epsilon | X^n]]/\epsilon.$$

Implying that:

$$\sup_{BL(1)} \left| \mathbb{E}[f(A_n)|X^n] - \mathbb{E}[f(B_n)|X^n] \right| \xrightarrow{P} 0,$$

as desired.<sup>15</sup>

*Q.E.D.*

---

<sup>15</sup>We are extremely thankful to an anonymous referee who suggested major simplifications to the previous version of the proof of this Lemma.



## A.4. Additional Lemmata

**LEMMA 5** (Dudley (2002), p. 395) *Let  $W_n^*$ ,  $Y_n^*$  be random variables dependent on the data  $X^n = (X_1, X_2, \dots, X_n)$  inducing the probability measures  $P_W^n$  and  $P_Y^n$  respectively. Let  $A \subset \mathbb{R}^k$  and let  $A^\delta = \{y \in \mathbb{R}^k : \|x - y\| < \delta \text{ for some } x \in A\}$ . Then,*

$$\begin{aligned} |P_W^n(A|X^n) - P_Y^n(A|X^n)| &\leq \frac{1}{\delta} \left| \mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n] \right| \\ &\quad + \min\{P_Y^n(A^\delta \setminus A|X^n), P_Y^n((A^c)^\delta \setminus A^c|X^n)\} \end{aligned}$$

PROOF: First observe that:

$$P_W^n(A|X^n) - P_Y^n(A|X^n) \leq P_W^n(A|X^n) - P_Y^n(A^\delta|X^n) + P_Y^n(A^\delta|X^n) - P_Y^n(A|X^n)$$

Define  $f(x) := \max(0, 1 - \|x - A\|/\delta)$ . Then,  $\delta f \in \text{BL}(1)$  and:

$$\begin{aligned} P_W^n(A|X^n) &= \int_A dP_W^n|X^n \\ &\leq \int f dP_W^n|X^n \\ &\quad (\text{since } f \text{ is nonnegative and } f(x) = 1 \text{ over } A) \\ &= \int_A dP_Y^n|X^n + \frac{1}{\delta} \left( \int_A \delta f dP_W^n|X^n - \int_A \delta f dP_Y^n|X^n \right) \\ &\leq \int f dP_Y^n|X^n + \frac{1}{\delta} \sup_{f \in \text{BL}(1)} \left| \mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n] \right| \\ &= \int_{A^\delta} f dP_Y^n|X^n + \frac{1}{\delta} \sup_{f \in \text{BL}(1)} \left| \mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n] \right| \\ &\leq P_Y^n(A^\delta|X^n) + \frac{1}{\delta} \sup_{f \in \text{BL}(1)} \left| \mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n] \right| \end{aligned}$$

It follows that:

$$P_W^n(A|X^n) - P_Y^n(A|X^n) \leq \frac{1}{\delta} \left| \mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n] \right| + (P_Y^n(A^\delta|X^n) - P_Y^n(A|X^n))$$

An analogous argument can be made for  $A^c$ . In this case we get:

$$P_W^n(A^c|X^n) - P_Y^n(A^c|X^n) \leq \frac{1}{\delta} \left| \mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n] \right| + (P_Y^n((A^c)^\delta|X^n) - P_Y^n(A^c|X^n)),$$

which implies that:

$$P_W^n(A|X^n) - P_Y^n(A|X^n) \geq -\frac{1}{\delta} \left| \mathbb{E}[f(W_n^*)|X^n] - \mathbb{E}[f(Y_n^*)|X^n] \right| - (P_Y^n((A^c)^\delta|X^n) - P_Y^n(A^c|X^n))$$

The desired result follows.

*Q.E.D.*