On the performance of weighted bootstrapped kernel deconvolution density estimators

Al-Sharadqah, Ali; Mojirsheibani, Majid; Pouliot, William

DOI:
10.1007/s00362-018-1006-0

On the performance of weighted bootstrapped kernel deconvolution density estimators

Ali Al-Sharadqah[1], Majid Mojirsheibani[23], and William Pouliot[4]
Department of Mathematics, California State University Northridge, CA, 91330, USA
and
Department of Economics, University of Birmingham, B15 2TT, UK

Abstract

We propose a weighted bootstrap approach that can improve on current methods to approximate the finite sample distribution of normalized maximal deviations of kernel deconvolution density estimators in the case of *ordinary smooth* errors. Using results from the approximation theory for weighted bootstrap empirical processes, we establish an unconditional weak limit theorem for the corresponding weighted bootstrap statistics. Because the proposed method uses weights that are not necessarily confined to be uniform (as in Efron's original bootstrap), it provides the practitioner with additional flexibility for choosing the weights. As an immediate consequence of our results, one can construct uniform confidence bands, or perform goodness-of-fit tests, for the underlying density. We have also carried out some numerical examples which show that, depending on the bootstrap weights chosen, the proposed method has the potential to perform better than the current procedures in the literature.

Keywords: Kernel, deconvolution, density, weighted bootstrap, CLT.

Mathematics Subject Classification: 62G07

1 Introduction

Consider the following deconvolution density estimation problem. Let $X_1, \ldots, X_n$ be independently and identically distributed (iid) observations from the convolution model

$$X_i = Z_i + \epsilon_i\,,\ i = 1, \ldots, n,$$

where $\epsilon_1, \ldots \epsilon_n$ are iid random variables with the known probability density function (pdf) ψ and the corresponding characteristic function $\phi_\psi(t) = \int_{\mathbb{R}} \psi(x)\, e^{itx} dx$; here $\epsilon_1, \ldots \epsilon_n$ are independent of the iid random variables $Z_1, \ldots, Z_n$. Then the kernel deconvolution density estimator of the pdf f of Z_1 at the point x is given by

$$f_n(x) \equiv f_{n,h}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx}\, \frac{\phi_K(ht) \cdot \widehat{\phi}_n(t)}{\phi_\psi(t)}\, dt\,, \tag{1}$$

[1]Email: ali.alsharadqah@csun.edu
[2]Corresponding author. Email: majid.mojirsheibani@csun.edu
[3]This work is supported by the NSF Grant DMS-1407400 of Majid Mojirsheibani.
[4]Email: w.pouliot@bham.ac.uk

where $\widehat{\phi}_n(t) = n^{-1}\sum_{j=1}^n e^{itX_j}$ is the empirical characteristic function of the X. Here ϕ_K is the characteristic function of the compactly supported kernel K, with the smoothing parameter $h > 0$, used in the construction of the estimator $f_{n,h}$ in (1). When the density f is $p \geq 0$ times differentiable then (1) can be extended to the deconvolution estimator of the ℓth derivative of f, $\ell = 0, 1, \ldots, p$, as follows

$$f_n^{(\ell)}(x) \equiv f_{n,h}^{(\ell)}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} (-it)^\ell\, e^{-itx}\frac{\phi_K(ht)\cdot\widehat{\phi}_n(t)}{\phi_\psi(t)}\,dt \;:=\; \frac{1}{nh^{\ell+1}}\sum_{j=1}^n L_n^{(\ell)}\left(\frac{x-X_j}{h}\right), \qquad (2)$$

where

$$L_n^{(\ell)}(x) := \frac{1}{2\pi}\int_{\mathbb{R}} (-it)^\ell\, e^{-itx}\frac{\phi_K(t)}{\phi_\psi(t/h)}\,dt\,, \quad \ell = 0, 1, \ldots, p. \qquad (3)$$

Here, $L_n^{(\ell)}(x)$ depends on n through $h \equiv h_n$. In passing we also note that the far right side of (2) has the appearance of the usual kernel density estimator. Such kernel-type deconvolution density estimators have been studied extensively in the literature. These include the work of Carroll and Hall (1988), Stefanski (1990), Stefanski and Carroll (1990), Zhang (1990), Fan (1991, 1992), Fan and Liu (1997), Wand (1998), Hesse (1999), Cator (2001), Delaigle and Gijbels (2004, 2006), Van Es and Uh (2004, 2005), Lacour (2006), Kulik (2008), Achilleos and Delaigle (2012), and Delaigle and Hall (2014). In a more recent result, Zamini et al. (2015) have studied kernel deconvolution density estimators in the context of multiplicative models.

Several authors have also developed density estimators for the heteroscedastic error model. Notable results along these lines include the work of Delaigle and Meister (2008) who introduce a kernel estimator of the density in the case of heteroscedastic errors, and establish consistency of their estimator. McIntyre and Stefanski (2011) propose a deconvolution density estimator with heteroscedastic errors that are normally distributed; these authors also study the integrated mean squared error of their estimator. Meister (2010) considers a special type of heteroscedasticity that corresponds to two types of contaminated data sets and establishes upper and lower bounds for the convergence rate of the proposed nonparametric density estimator. Another relevant results along these lines is that of Chesneau and Fadili (2013) who propose a wavelet-based density estimator in the heteroscedastic model and study its mean integrated squared error.

To address the limiting distribution of kernel deconvolution density estimators (properly normalized), one must carefully take into account the rate at which the modulus of the characteristic function of the error, $\phi_\psi(t)$, decays to zero as t diverges. More specifically two cases are to be distinguished: the *supersmooth* case where $|\phi_\psi(t)|$ decays to zero at an exponential rate, and the *ordinary smooth* case where it decays at a polynomial rate. In the supersmooth case, the asymptotic normality of $f_n(x)$ was studied and established by Zhang (1990), Fan (1991), and van Es and Uh (2005). Also see Masry (1993) who established similar results for dependent (stationary) sequences. For the ordinary smooth case, asymptotic normality was first established by Fan (1991); also see Fan and Liu (1997) as well as van Es and Kok (1998). The work of Fan (1991) also includes the asymptotic normality of the derivative estimators $f_n^{(\ell)}(x)$ of $f^{(\ell)}(x)$. Such results can be used to form pointwise confidence intervals or to carry out tests of significance. On the other

hand, to construct uniform confidence bands for f, one needs to derive the limiting distribution of the properly normalized maximal deviations of $f_n(x)$ from $f(x)$ (or from $E(f_n(x))$). For the case of supersmooth errors, van Es and Gugushvili (2008) were the first to derive this limiting distribution.

For the ordinary smooth case, which is also the focus of our work here (from a weighted bootstrap approximation point of view), the limiting distribution of the properly normalized versions of the statistics $\sup_{t\in[0,1]} \left|f_n^{(\ell)}(t) - f^{(\ell)}(t)\right| / \sqrt{g(t)}\,,\ \ell = 0, \ldots, p$, where g is the density function of the random variable X, has been studied by Bissantz et al. (2007). Of course, g is unknown in practice and will be replaced by some suitable density estimator. These results can be used to construct uniform confidence bands for f. However, because of the poor coverage properties of the resulting bands, Bissantz et al. (2007) use Efron's (1979) original bootstrap algorithm to construct percentile bootstrap bands for f. In this paper we propose a weighted bootstrap approach that can be viewed as a generalization of Efron's (1979) method. We also note that Efron's bootstrap is itself a weighted bootstrap, where the weights form an n-category multinomial random vector with probabilities $(n^{-1}, \ldots, n^{-1})$. The potential applications of the weighted bootstrap have been demonstrated by several authors. For example, in the case of confidence intervals for a mean, it has been shown in Shao and Tu (1995; pp 440-441) that weighted bootstrap with random weights has a better coverage probability than Efron's bootstrap. This is particularly true with smaller sample sizes. Our numerical results in this paper confirm some of these findings and show that the proposed weighted bootstrap has the ability to improve on the performance of current methods in deconvolution density estimation in the ordinary smooth case. Furthermore, it has been shown (see for example Burke (2000), Hall and Mammen (1994), Hórvath et al. (2000), Chiang et al. (2005), and Chiang et al. (2009)) that in some applications the weighted bootstrap can be computationally more efficient than Efron's (1979) original algorithm. In Section 2.2 we provide more detailed comments about various classes of weights employed in the literature as well as the types of weights proposed in this paper.

From an applied point of view, our proposed approximation methods have the potential to provide flexible tools in the statistical analysis of data arising from measurement errors models in a variety of fields such as economics, imaging and signal processing, biometrics, genetics, and medicine. Here we discuss a few such applications:

(i) One area of genetics looks at how mutations affect organismal fitness. To do this requires estimating the mutation effect distribution because this allows one to assess the frequency of neutral or nearly neutral effects of mutations on their fitness. Estimating the distribution of the mutation effect is difficult because the effect of organismal fitness is small and subject to measurement errors. Traditionally, the density of this distribution was assumed to follow some parametric form and then the parameters associated with it were estimated. In practice, the parametric family was not checked for validity. This led Lee et al. (2010) to propose a nonparametric density estimator based on deconvolution approaches. Using virus lineage data taken from Burch et al. (2007), they estimate the mutation effect distribution to provide an estimate of the frequency of neutral or nearly neutral mutations. It is always desirable to be able to study the uniform behavior of the

underlying density.

(ii) Errors-in-variables is a popular modeling device in economics because many of the variables described by theory are unobserved. They, therefore, cannot be used in regression models to test the validity of economic theory. In particular, according to one model of a firm's demand for investment, the correct measure of this demand is the marginal Tobin's q. Marginal q is unobserved because it is defined as the ratio of expected cash flow from investing to the accounting value of a firm's assets. Because one cannot observe expected cash flow, this led economists to find proxies for it. One proxy is called average Tobin's q. The relationship between average and marginal q is modeled as errors-in-variables (see Galvao et al. (2018) for more details). In view of our setting (i.e., $X = Z + \epsilon$), here Z would represent marginal Tobin's q and X would represent average Tobin's q. As the former determines a firm's investment demand, it is useful to have the knowledge of the density of marginal Tobin's q, a variable that is unobserved.

(iii) Other applications appear in computer vision as well as image and signal processing. These applications are discussed in some detail in Meister (2009). Signals and images are usually affected by some noise because of their surrounding environment, leading to the detection of blurry images and signals (Benšić and Sabo (2016)). Therefore, a deconvolution approach is aimed at recovering better signals and images from their noisy versions. This technique is known as de-noising or de-blurring. Applications of deconvolution methods in signal processing with connection to spectrometry and chemistry can be found in Cornelis and Hassellöv (2014).

2 Main results

2.1 Preliminaries

This subsection is devoted to the presentation of a number of standard assumptions used in the literature (and in this paper) that were also used by Bissantz et al. (2007) to establish the asymptotic distribution of the normalized versions of the maximal deviations of the kernel deconvolution estimator $f_n^{(\ell)}(t)$. More specifically, to summarize the existing results, consider the ordinary smooth case assumption:

Assumption A0.
$\phi_\psi(t) \cdot t^\beta \to C_0$, as $t \to \infty$, for some $\beta \geq 0$ and $C_0 \in \mathbb{C}\backslash\{0\}$.

When the above assumption holds, the results of Fan (1991) and Fan and Liu (1997) show that

$$h^\beta L_n^{(\ell)}(x) \longrightarrow L^{(\ell)}(x)\,, \text{ as } h \to 0\,,$$

where, for $\ell = 0, \ldots, p$,

$$L^{(\ell)}(x) = \frac{1}{2\pi C_0} \int_0^\infty (-it)^\ell \, e^{-itx} t^\beta \phi_K(t)\, dt + \frac{1}{2\pi \overline{C}_0} \int_{-\infty}^0 (-it)^\ell \, e^{-itx} |t|^\beta \phi_K(t)\, dt\,, \tag{4}$$

where $\overline{C}_0$ is the complex conjugate of C_0.

Additionally, the following assumptions are the same as those used by Bissantz et al. (2007):

Assumption A1.
The function $\phi_K(t) = \int_{\mathbb{R}} K(x)\, e^{itx} dx$ is symmetric, three-times differentiable and supported on $[-1, 1]$ and satisfies $|\phi_K(t)| = 1$ for $|t| \leq c$, for some $c \in (0, 1)$, and $|\phi_K(t)| \leq 1$ for $|t| > c$.

Assumption A2.

(a) $\int_{\{|x|>e\}} |L_n^{(\ell+1)}(x)|\, |x|^{3/2} [\log\log|x|]^{1/2} dx = \mathcal{O}(h^{-\beta})$, where β is as in assumption A0.

(b) $\int_{\{|x|>e\}} |h^{\beta} L_n^{(\ell+1)}(x) - L^{(\ell+1)}(x)|\, |x|^{1/2} [\log\log|x|]^{1/2} dx = \mathcal{O}(h^{\gamma+1/2})$, for some $\gamma > 0$, where $L^{(\ell)}$ is as in (4).

Assumption A3.

(a) The density g of X is bounded, and bounded away from zero, on $[0, 1]$. Additionally, $g^{1/2}$ has a bounded derivative.

(b) The characteristic function $\phi_f(t) = \int_{\mathbb{R}} f(x)\, e^{itx} dx$ of f satisfies $\int_{\mathbb{R}} |\phi_f(t)|\, |t|^{r-1} dt < \infty$ for some $r > p + 1$, where, as before, $p \geq 0$ is the number of times f is differentiable.

Now consider the statistic

$$\widehat{\Gamma}_n(t, \ell) = \frac{n^{1/2} h^{\beta+\ell+1/2}}{\sqrt{\widehat{g}_n(t)}} \left(f_n^{(\ell)}(t) - f^{(\ell)}(t) \right), \quad t \in [0, 1], \quad \ell = 0, \ldots, p, \tag{5}$$

where $\widehat{g}_n$ is an estimate of the density g. Bissantz et al. (2007) have established the limiting distribution of the properly normalized supremum of $|\widehat{\Gamma}_n(t, \ell)|$ under the assumption that $\widehat{g}_n$ satisfies

$$\sup_{0 \leq t \leq 1} |\widehat{g}_n(t) - g(t)| = o_p\left(\left(\log(1/h) \right)^{-1} \right), \quad \text{as } n \to \infty \text{ (and thus } h \to 0). \tag{6}$$

More specifically, for $\ell = 0, \ldots, p$, if we put

$$C_{1,\ell} = (2\pi |C_0|^2)^{-1} \int_{\mathbb{R}} x^{2(\beta+\ell)} \phi_K^2(x) dx, \quad C_{2,\ell} = \frac{\int_{\mathbb{R}} x^{2(\beta+\ell+1)} \phi_K^2(x) dx}{\int_{\mathbb{R}} x^{2(\beta+\ell)} \phi_K^2(x) dx} \tag{7}$$

$$d_{n,\ell} = (2\log(1/h))^{1/2} + (2\log(1/h))^{-1/2} \log\{C_{2,\ell}^{1/2}/(2\pi)\}, \tag{8}$$

then the following result is an immediate consequence of the work of Bissantz et al. (2007), (see their Theorem 1 in conjunction with their Corollary 1):

Theorem 1 *Let $\widehat{\Gamma}_n(t, \ell)$ be as in (5). Suppose that assumptions A0-A3 hold and that*

$$nh^{2(\beta+r)-1} \log(1/h) \to 0 \quad \textit{and} \quad nh^{2(\beta+\ell)+1} / \log(1/h) \to \infty,$$

as $n \to \infty$. If the density estimator $\widehat{g}_n$ satisfies (6) then, as $n \to \infty$,

$$P\left\{ \sqrt{2\log(1/h)} \left(C_{1,\ell}^{-1/2} \sup_{0 \leq t \leq 1} \left| \widehat{\Gamma}_n(t, \ell) \right| - d_{n,\ell} \right) \leq z \right\} \to \exp\left(-2e^{-z}\right), \quad \ell = 0, \ldots, p. \tag{9}$$

The above result can be used to construct $(1-\alpha)100\%$ uniform asymptotic confidence bands for $f^{(\ell)}(t)$, given by

$$\mathcal{I}_{n,\ell}(\alpha) := f_n^{(\ell)}(t) \pm \sqrt{\frac{\widehat{g}_n(t)\cdot C_{1,\ell}}{nh^{2(\beta+\ell)+1}}}\left(\frac{x^{(\alpha)}}{\sqrt{2\log(1/h)}} + d_{n,\ell}\right), \quad 0 \le t \le 1\,,$$

where $x^{(\alpha)} = -[\log\log(\frac{1}{1-\alpha}) - \log 2]$ solves the equation $\exp(-2\exp(-x)) = 1-\alpha$. In other words, $P\{f^{(\ell)}(t) \in \mathcal{I}_{n,\ell}(\alpha),\ 0 \le t \le 1\} \to 1-\alpha$. In passing, we point out that Bissantz et al. (2007) have also considered the following version of $\widehat{\Gamma}_n(t,\ell)$ in (5):

$$\Gamma_n(t,\ell) = \frac{n^{1/2}h^{\beta+\ell+1/2}}{\sqrt{g(t)}}\left(f_n^{(\ell)}(t) - E\big[f_n^{(\ell)}(t)\big]\right), \quad t \in [0,1]\,, \quad \ell = 0,\ldots,p\,. \tag{10}$$

Another closely related statistic is given by

$$\widetilde{\Gamma}_n(t,\ell) = \frac{n^{1/2}h^{\beta+\ell+1/2}}{\sqrt{g(t)}}\left(f_n^{(\ell)}(t) - f^{(\ell)}(t)\right), \quad t \in [0,1]\,, \quad \ell = 0,\ldots,p\,. \tag{11}$$

Here, both (10) and (11) have the same limiting distribution as the one in (9); however, clearly (5) is more appropriate for constructing uniform confidence intervals for $f^{(\ell)}(t)$. Unfortunately, approximate confidence bands based on (9) can have poor coverage probabilities; this will be taken up in the next section where we propose weighted bootstrap approximations to approximate the distribution of $\widehat{\Gamma}_n(t,\ell)$.

2.2 The proposed weighted bootstrap

Given the typically slow rates of convergence (logarithmic) of the normalized maximal deviations of nonparametric density estimators (see, for example, Hall (1991) and Konakov and Piterbarg (1984) for the case of kernel density estimators), Bissantz et al. (2007) proposed percentile type bootstrap confidence bands based on Efron's (1979) original bootstrap. Their approach replaces $\Gamma_n(t,\ell)$ in (10) by

$$\Gamma_n^*(t,\ell) = \frac{n^{1/2}h^{\beta+\ell+1/2}}{\sqrt{\widehat{g}_n(t)}}\left(f_n^{*(\ell)}(t) - f_n^{(\ell)}(t)\right), \quad t \in [0,1]\,, \quad \ell = 0,\ldots,p\,, \tag{12}$$

where

$$f_n^{*(\ell)}(t) = \frac{1}{nh^{\ell+1}}\sum_{j=1}^{n} L_n^{(\ell)}\left(\frac{x - X_j^*}{h}\right), \tag{13}$$

with $X_1^*,\ldots,X_n^*$ representing a sample of size n drawn, with replacement, from $X_1,\ldots,X_n$. Simulating $\Gamma_n^*(t,\ell)$ a large number of times, Bissantz et al. (2007) used the bootstrap $(1-\alpha)$-quantile, $q_{1-\alpha}^*$, of $\Gamma_n^*(t,\ell)$ to construct the bootstrap percentile-type confidence bands for $f^{(\ell)}(t)$, given by $f_n^{(\ell)}(t) \pm n^{-1/2}n^{-(\beta+\ell+1/2)}\, q_{1-\alpha}^*\sqrt{\widehat{g}_n(t)}\,,\ \ t \in [0,1]\,,\ \ell = 0,\ldots,p$.

Several authors have proposed the weighted bootstrap in the literature as a generalization of Efron's (1979) original bootstrap. To the best of our knowledge, the first paper that used the concept of weighted bootstrap, with weights different from those of Efron, is that of Rubin (1981). Rubin uses the n spacings of $n-1$ ordered uniform $(0,1)$ random variables for the weights; this is equivalent to the Dirichlet weights $(W_{n,1},\ldots,W_{n,n}) =_d$ Dirichlet$_n(1,1,\ldots,1)$. The case where $(W_{n,1},\ldots,W_{n,n}) =_d$ Dirichlet$_n(4,4,\ldots,4)$ was studied by Weng (1989), and Zheng and Tu (1988, Remark 5). The random weighting method employed by Gao and Zhong (2010) in the problem of kernel density estimation also uses Dirichlet$_n(1,1,\ldots,1)$ weights. We also note that if $E_1,\ldots,E_n$ are iid Exp(1) random variables then the random vector of weights $(\frac{E_1}{\sum_i E_i},\ldots,\frac{E_n}{\sum_i E_i})$ is also Dirichlet$_n(1,1,\ldots,1)$. In fact, some authors have considered general weights of the form $(\frac{V_1}{\sum_i V_i},\ldots,\frac{V_n}{\sum_i V_i})$, where V_i's are iid positive random variables with finite mean and variance; see, for example, Mason and Newton (1992, Ex. 2.1). These different weights lead to a general approach to study the class of exchangeable bootstrap weights, originally studied by Mason and Newton (1992) and subsequently by Praestgaard and Wellner (1993), Janssen and Pauls (2003), and Janssen (2005) among others. An interesting application of exchangeable weighted bootstrap appears in the work of Bouzebda et al. (2017).

A different class of bootstrap weights, which is the subject of our present work in this paper, is that of Horváth et al. (2000), Horváth (2000), Burke (1998, 2000), and their more powerful generalizations in Burke (2010). What makes these weights particularly more appropriate for our work is their flexibility and, more importantly, the fact that the corresponding weighted bootstrap empirical processes can be approximated by a sequence of Brownian bridges with the best possible rate. This result, which is due to Horváth et al. (2000), and its mighty generalization to multivariate empirical processes, due to Burke (2010), make this class of bootstrap weights quite suitable for approximating the distribution of many complicated statistics. In fact, in this paper we propose a weighted bootstrap approach, based on this latter class of weights, to approximate the limiting distribution in (9) for the statistics given by (5) and (11). Recent results along these lines include the work of Mojirsheibani and Pouliot (2017) on a weighted bootstrap approximation of the L_p norms of kernel density estimators in two-sample problems, the results of Ahlgren and Catani (2017) on tests for autocorrelation, the work of Kojadinovic and Yan (2012) as well as Kojadinovic, Yan, and Holmes (2011) on the applications of the weighted bootstrap to goodness-of-fit tests. A rather thorough discussion of different types of weighted bootstrap procedures (conditional and unconditional) can be found in Cheng and Huang (2010) and Kosorok (2008; Sec. 10). The monograph by Barbe and Bertail (1995) provides a survey of many other results on weighted bootstrap. It also seems probable that more sophisticated weighted bootstrap methods would be equally good or better than the ones discussed here, but we have not been able to show this thus far.

Our method works as follows. Let $\xi_1,\ldots,\xi_n$ be iid random variables, independent of the data $X_1,\ldots,X_n$, with mean $E(\xi_1)=\mu$ and Var$(\xi_1)=1$ and consider the following weighted bootstrap version of $f_n^{(\ell)}(t)$

$$f_{nn}^{(\ell)}(t) = (nh^{\ell+1})^{-1}\sum_{j=1}^{n}(1+\xi_j-\bar{\xi})\,L_n^{(\ell)}\left(\frac{x-X_j}{h}\right), \tag{14}$$

where $L_n^{(\ell)}(x)$ is given by (3) and $\bar{\xi}$ is the sample mean of $\xi_1, \ldots, \xi_n$. Observe that if $(1+\xi_i-\bar{\xi})$ is replaced by $W_{n,i}$ in (14), where $(W_{n,1}, \ldots, W_{n,n})$ is an n-category multinomial random vector with probabilities $(\frac{1}{n}, \ldots, \frac{1}{n})$, then $f_{nn}^{(\ell)}(t)$ coincides with $f_n^{*(\ell)}(t)$ in (13). Next, consider the bootstrap statistics

$$\widetilde{\Gamma}_{nn}(t,\ell) = \frac{n^{1/2}h^{\beta+\ell+1/2}}{\sqrt{\widehat{g}_n(t)}} \left(f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t)\right), \quad t \in [0,1], \quad \ell = 0, \ldots, p, \tag{15}$$

and

$$\widehat{\Gamma}_{nn}(t,\ell) = \frac{n^{1/2}h^{\beta+\ell+1/2}}{\sqrt{\widehat{g}_{nn}(t)}} \left(f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t)\right), \quad t \in [0,1], \quad \ell = 0, \ldots, p, \tag{16}$$

where $\widehat{g}_n$ is an estimator of the density g of X_i's and $\widehat{g}_{nn}$ in (16) is the bootstrap version of $\widehat{g}_n$. We note that (15) is the weighted bootstrap counterpart of (11), whereas (16) is the counterpart of (5). As for the estimator $\widehat{g}_n$ that appears in (15), we follow Bissantz et al. (2007) and consider any estimator that satisfies (6). However, we also consider the specific choice where $\widehat{g}_n$ is a kernel density estimator:

$$\widehat{g}_n(x) = (n\lambda_n)^{-1} \sum_{i=1}^{n} \mathcal{H}\big((x - X_i)/\lambda_n\big), \tag{17}$$

where $\mathcal{H} : \mathbb{R} \to \mathbb{R}_+$ is the kernel used with the smoothing parameter $\lambda_n (\to 0, \text{ as } n \to \infty)$. The bootstrap estimator $\widehat{g}_{nn}$ in (16) is given by

$$\widehat{g}_{nn}(x) = (n\lambda_n)^{-1} \sum_{i=1}^{n} (1 + \xi_i - \bar{\xi})\, \mathcal{H}\big((x - X_i)/\lambda_n\big), \tag{18}$$

where $\xi_1, \ldots, \xi_n$ are as in (14). As for the choice of the kernel $\mathcal{H}$, we require:

Assumption (H). The kernel $\mathcal{H}$ is nonnegative, symmetric about zero, vanishes outside an interval $[-A, A]$, and is absolutely continuous on $[-A, A]$. The derivative $\mathcal{H}'$ exists and satisfies $\int_{\{|x|>e\}} |x|^{3/2} (\log\log|x|)^{1/2} |\mathcal{H}'(x)| dx < \infty$. Furthermore, $\mathcal{H}$ satisfies $\int \mathcal{H}(x)dx = 1$.

To state our main results, we first state an assumption regarding the choice of the random variables $\xi_1, \ldots, \xi_n$.

Assumption A4.
The random variables $\xi_1, \ldots, \xi_n$ are iid with mean μ and variance 1, and are chosen independent of the data $X_1, \ldots, X_n$. Furthermore, ξ_1 has a finite moment generating function in an open interval around the origin.

Theorem 2 *Let $\widetilde{\Gamma}_{nn}(t,\ell)$ and $\widehat{\Gamma}_{nn}(t,\ell)$ be the weighted bootstrap statistics in (15) and (16), respectively. Also, let $C_{1,\ell}$ and $d_{n,\ell}$ be as in (7) and (8). Suppose that assumptions A0-A4 are satisfied and that $nh^{2(\beta+r)-1}\log(1/h) \to 0$ and $nh^{2(\beta+\ell)+1}/\log(1/h) \to \infty$, as $n \to \infty$.*

(i) *If $\widehat{g}_n$ is any estimator satisfying (6) then, as $n \to \infty$,*

$$P\left\{\sqrt{2\log(1/h)}\left(C_{1,\ell}^{-1/2} \sup_{0\le t\le 1}\left|\widetilde{\Gamma}_{nn}(t,\ell)\right| - d_{n,\ell}\right) \le z\right\} \to \exp\left(-2e^{-z}\right), \quad \ell = 0, \ldots, p.$$

(ii) *Let $\widehat{g}_{nn}$ be as in (18), where $\lambda_n = n^{-\delta}$, $\delta \in (\frac{1}{5}, \frac{1}{2})$, and suppose that assumption (H) is satisfied. If $\log(1/h)\sqrt{\log n/(n\lambda_n)} \to 0$, as $n \to \infty$, then*

$$P\left\{\sqrt{2\log(1/h)}\left(C_{1,\ell}^{-1/2}\sup_{0\le t\le 1}\left|\widehat{\Gamma}_{nn}(t,\ell)\right| - d_{n,\ell}\right) \le z\right\} \to \exp\left(-2e^{-z}\right)\ , \quad \ell = 0,\ldots,p\,.$$

Remark 1 Theorem 2 shows that the bootstrap statistics $\widetilde{\Gamma}_{nn}$ and $\widehat{\Gamma}_{nn}$ both yield the same correct limiting distribution. However, from a practical point of view, in most cases one should employ $\widehat{\Gamma}_{nn}(t,\ell)$. This is because the density g of X is virtually always unknown and, as a result, one has to work with the statistic (5) instead of (11). Consequently, with g unknown, the statistic (16) (and not (15) or (12)) is the right bootstrap counterpart for the statistic in (5). To better appreciate this, we give an analogy between our statistics and some classical results: it is well known (see, for example, Hall (1992; Sec. 2.4)) that, under the *smooth function of means* model, and the existence of sufficient moments, the asymptotically normal statistic $T_n = \sqrt{n}(\widehat{\theta} - \theta)/\widehat{\sigma}$ admits standard Edgeworth expansions whenever the characteristic function of the underlying distribution obeys the so-called *Cramér's continuity* condition; here $\widehat{\sigma}^2$ is a consistent estimator of the asymptotic variance σ^2 of $n^{1/2}\widehat{\theta}$. Using arguments based on such expansions, it can be shown (see, for example, Hall (1992; Sec. 3.3) and Singh (1981)) that there is significant gain in employing the bootstrap if one uses $T_n^* = \sqrt{n}(\widehat{\theta}^* - \widehat{\theta})/\widehat{\sigma}^*$ as the bootstrap counterpart of T_n, where $\widehat{\theta}^*$ and $\widehat{\sigma}^*$ are the bootstrap versions of $\widehat{\theta}$ and $\widehat{\sigma}$. This gain can be expressed in terms of the size of the approximation error, $|P\{T_n \le x\} - P^*\{T_n^* \le x\}| = O_p(n^{-1})$, which holds uniformly in x, whereas the classical CLT only yields $|P\{T_n \le x\} - \Phi(x)| = O(n^{-1/2})$. Here P^* is the bootstrap probability. On the other hand, the results in Hall (1992; Sec. 3.3) also show that if one uses the wrong bootstrap statistic $S_n^* = \sqrt{n}(\widehat{\theta}^* - \widehat{\theta})/\widehat{\sigma}$ (instead of T_n^*) to approximate T_n then, in general, one has $|P\{T_n \le x\} - P^*\{S_n^* \le x\}| = O_p(n^{-1/2})$, which is not any better than the classical CLT. This analogy shows that the correct bootstrap counterpart of (5) is given by (16). Using the weighted bootstrap statistic (15) instead of (16) to approximate (5) is, in a sense, equivalent to looking up the Z table instead of the t table.

Bootstrap weights

According to assumption A4 and Theorem 2, the proposed bootstrap approximations are asymptotically valid provided that the bootstrap weights are chosen to have a finite moment generating function (mgf) and unit variance. Furthermore, the results of Burke (2010) imply that the assumption of a finite mgf can be further relaxed to one based on the existence of sufficient moments. This implies that in practice there is a great deal of flexibility in choosing bootstrap weights. On the other hand, in a given application, the finite-sample performance of some weights can be better than that of others. As a matter of fact, Barbe and Bertail (1995; Ch. II) point out that *the choice of the adequate weights depends on the priorities of the statistician*. For example, weights that yield good coverage probabilities for confidence intervals in a specific problem are not necessarily the same as weights that yield good approximations of the distribution of the corresponding statistics in the same problem. This type of scenario is not particular to weighted bootstrap and,

in fact, appears in certain other areas of probability and statistics. An analogy is the choice of the kernel in nonparametric kernel regression (or density) estimation, where various asymptotic results can be established provided that the chosen kernel satisfies certain assumptions (such as boundedness, smoothness, absolute integrability, etc.) But, there are many different kernels that satisfy such conditions, and although some exhibit better finite-sample performance in a given statistical problem, there are no rules for choosing one. In fact, many practitioners simply use popular kernels such as Gaussian, Uniform, or Epanechnikov.

In the case of bootstrap weights that appear in the proposed approach, a popular choice in the literature appears to be the Gaussian weight; see, for example, Burke (1998, 2000), Horváth et al. (2000), Kojadinovic and Yan (2012), and Mojirsheibani and Pouliot (2017).

Numerical results

In what follows, we provide some numerical examples in order to assess the finite-sample performance of the following methods and their applications to the construction of confidence bands: the asymptotic theory (Theorem 1), the percentile type bootstrap of Bissantz et al. (2007) based on 13), and the proposed weighted bootstrap statistics (Theorem 2). Additionally, based on the recommendations of an anonymous referee, we have also considered the *random weighting* method, as described in Section 2.2, where we have followed Gao and Zhong (2010) and employed the popular Dirichlet$_n(1, 1, \ldots, 1)$ random weights. Our examples show that the proposed weighted bootstrap can perform quite well in terms of approximating the finite-sample distribution of $\widehat{\Gamma}_n(t, \ell)$ in (5) and that of $\widetilde{\Gamma}_n(t, \ell)$ in (11). In what follows, we consider random samples of sizes $n =$ 50, 100, and 500 drawn from the distribution (density) f of Z. As for the choice of f, we consider two different distributions: $N(0.5,\, 0.2^2)$, i.e., a normal distribution with mean 0.5 and standard deviation 0.2, and a gamma distribution with the shape parameter 5 and scale parameter 0.1.

Here, the distribution of the measurement error ϵ is taken to be Laplacian with the pdf $\psi(x) = \frac{1}{2\sigma}\exp(-|x|/\sigma)$, $-\infty < x < \infty$, where $\sigma = 0.1$. The characteristic function of this Laplace distribution is $\phi_\psi(t) = (1+\sigma^2 t^2)^{-1}$, which yields $\beta = 2$ and $C_0 = \sigma^{-2}$ in assumption A0. Furthermore, it is straightforward to see that for a Laplace error distribution, the function $L_n(x) \equiv L_n^{(0)}(x)$ in (3) reduces to $L_n(x) = K(x) - (\sigma^2/h^2)K''(x)$, where K is the kernel used with the smoothing parameter h. Therefore, the density estimate f_n in (1) can be expressed in closed form. As for the choice of the kernel K with a compactly supported Fourier transform $\phi_K(t)$, to be used in the construction of the estimate of the density f of Z, we consider the flat Fourier transform $\phi_K(t) = I\{-1 \le t \le 1\}$ corresponding to the kernel $K(x)$ which is proportional to $\sin(x)/x$. To estimate the density g of X, based on $X_1, \ldots, X_n$, we employ the kernel density estimator $\widehat{g}_n(x)$ in (17) where $\mathcal{H}$ is a Gaussian kernel, truncated at -4 and 4 (which satisfies the conditions of assumption (H)). The smoothing parameter λ_n of $\mathcal{H}$ was selected via cross-validation. The selection of the bandwidth h for the density estimator f_n is more subtle. Here we have used two estimators: the first one is the bandwidth that minimizes the approximate MISE, where we have used the plug-in method of Wang and Wang (2011) which is available in the R package called "decon"; this estimator will be denoted by $H1$ throughout this section. Our second estimator of h is based on the minimization of $\sup_t |f_n(t) - f(t)|$, which was also discussed by

Bissantz et al. (2007) and will be denoted by $H2$ in this paper. Of course, in practice, $H2$ is not available because the density f is unknown. However, since we are merely comparing various estimators, regardless of whether a choice of h is good or poor for a particular data set, it will be good or poor for all the methods discussed here (because we are using the same density estimator).

To implement the proposed weighted bootstrap, we consider two choices for the distribution of the weights $\xi_1, \ldots, \xi_n$: the standard normal, N(0, 1), and the standard exponential, Exp(1). Next, to assess the performance of various approximations, we computed the statistic

$$Y_n \;:=\; \sqrt{2\log(1/h)}\left\{C_{1,0}^{-1/2}\sup_{0\le t\le 1}\left|\widehat{\Gamma}_n(t,0)\right| - d_{n,0}\right\},$$

for each of the two bandwidth estimators (of f_n) as well as each sample size $n = 50, 100, 500$, where $\widehat{\Gamma}_n(t,0)$, $C_{1,0}$, and $d_{n,0}$ are as in (5), (7), and (8), respectively. In addition to the statistics Y_n above, we also computed $B = 1000$ copies of the following bootstrap statistic for each of the two bandwidth estimators and each sample size n:

$$\begin{aligned}
Y_n^* &:= \sqrt{2\log(1/h)}\left\{C_{1,0}^{-1/2}\sup_{0\le t\le 1}|\Gamma_n^*(t,0)| - d_{n,0}\right\}\\
\widetilde{Y}_{nn} &:= \sqrt{2\log(1/h)}\left\{C_{1,0}^{-1/2}\sup_{0\le t\le 1}\left|\widetilde{\Gamma}_{nn}(t,0)\right| - d_{n,0}\right\}\\
\widehat{Y}_{nn} &:= \sqrt{2\log(1/h)}\left\{C_{1,0}^{-1/2}\sup_{0\le t\le 1}\left|\widehat{\Gamma}_{nn}(t,0)\right| - d_{n,0}\right\}\\
Y_{nn}^\dagger &:= \sqrt{2\log(1/h)}\left\{C_{1,0}^{-1/2}\sup_{0\le t\le 1}\left|\Gamma_{nn}^\dagger(t,0)\right| - d_{n,0}\right\},
\end{aligned}$$

where $\Gamma_n^*(t,0)$, $\widetilde{\Gamma}_{nn}(t,0)$, and $\widehat{\Gamma}_{nn}(t,0)$ are as in (12), (15), and (16), respectively, and $\Gamma_{nn}^\dagger(t,0)$ is as in (16) but with bootstrap density estimators constructed based on Dirichlet$_n(1,1,\ldots,1)$ weights (as in Gao and Zhong (2010)). The computation of the weighted bootstrap statistics $\widetilde{Y}_{nn}$ and $\widehat{Y}_{nn}$ was carried out for both weight distributions, N(0,1) and Exp(1). In practice, to compute the supremum functionals in all the above expressions, we used the maximum of the corresponding function over a grid of 200 equally-spaced values of t in the $[0\,,1]$. Changing the grid size to as large as 500 did not make any noticeable changes. To present our numerical results, first observe that if n is not "very small" then by Theorem 1 the random variable $U := \exp\{-2\exp(-Y_n)\}$ has an approximate Unif$\,[0,1]$ distribution. Similarly, if the bootstrap methods used are good approximations, the random variables $U^* := B^{-1}\sum_{j=1}^B I\{Y_{n,j}^* \le Y_n\}$, $\widetilde{U} := B^{-1}\sum_{j=1}^B I\{\widetilde{Y}_{nn,j} \le Y_n\}$, $\widehat{U} := B^{-1}\sum_{j=1}^B I\{\widehat{Y}_{nn,j} \le Y_n\}$, and $U^\dagger := B^{-1}\sum_{j=1}^B I\{Y_{nn,j}^\dagger \le Y_n\}$ will each have an approximate Unif$\,[0,1]$ distribution, where $Y_{n,j}^*$ is the jth copy of Y_n^*, based on the jth bootstrap sample, $j = 1,\ldots,B$, (similarly, $\widetilde{Y}_{nn,j}$, $\widehat{Y}_{nn,j}$, and $Y_{nn,j}^\dagger$ are the jth copies of the statistics $\widetilde{Y}_{nn}$, $\widehat{Y}_{nn}$, and $Y_{nn}^\dagger$ based on the jth set of weights). Here, $B = 1000$ as before. Repeating the entire above process a total of 300 times yields $U_1,\ldots,U_{300}$, $U_1^*,\ldots,U_{300}^*$, $\widetilde{U}_1,\ldots,\widetilde{U}_{300}$, $\widehat{U}_1,\ldots,\widehat{U}_{300}$, and $U_1^\dagger,\ldots,U_{300}^\dagger$ for each setup (corresponding to the choices for the distribution of f, the two estimates of the bandwidth h, the two choices for the distribution of the weights for weighted

bootstraps (normal and exponential), and the three sample sizes $n = 50$, 100, 500). In Figure 1 we have plotted the empirical distribution function of $U_1, \ldots, U_{300}$, $U_1^*, \ldots, U_{300}^*$, $\widetilde{U}_1, \ldots, \widetilde{U}_{300}$, $\widehat{U}_1, \cdots, \widehat{U}_{300}$, and $U_1^\dagger, \ldots, U_{300}^\dagger$ for the case where f is the density of a Gamma distribution with the shape parameter 5 and scale parameter 0.1; we have also included the 45° line (which is the true CDF of a Unif $[0, 1]$ random variable).

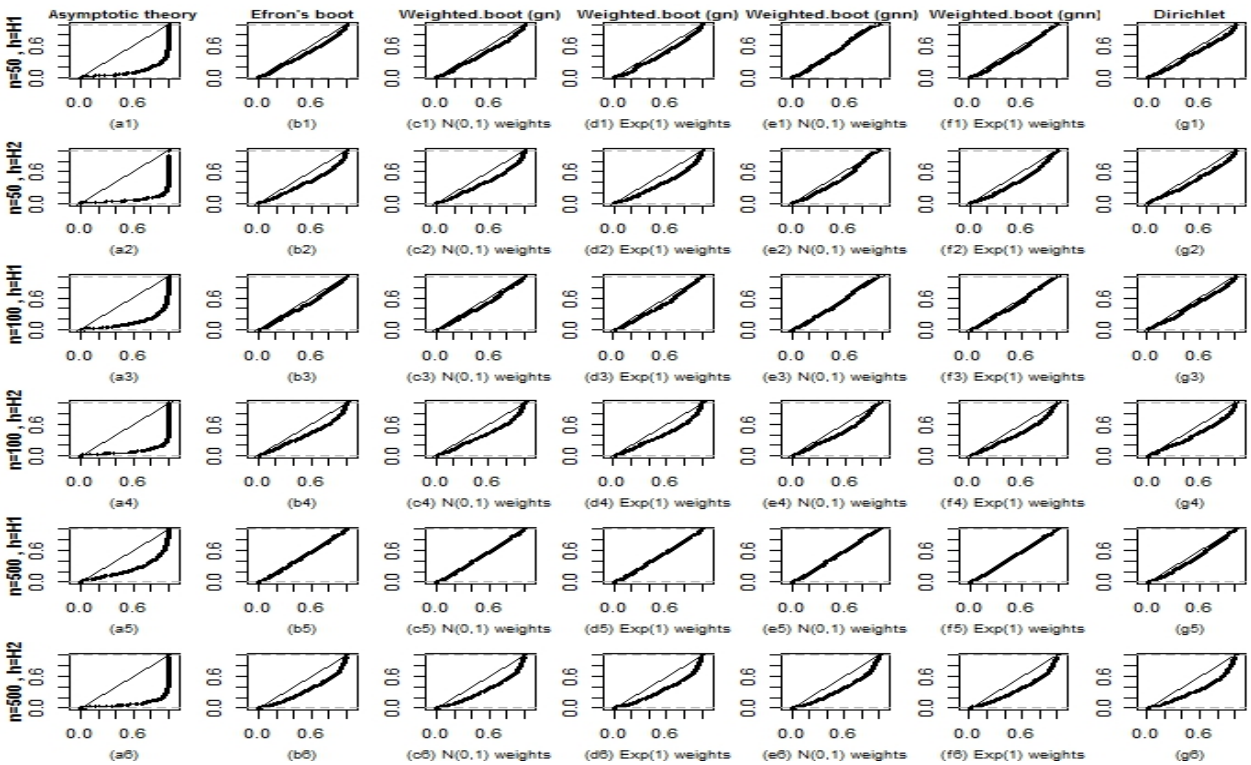

Figure 1: Plots of empirical cdf's when f is the density of a Gamma distribution with the shape parameter 5 and scale parameter 0.1. Here, plots (a1) to (a6) correspond to $U_1, \cdots, U_{300}$, (b1) to (b6) correspond to $U_1^*, \cdots, U_{300}^*$, (c1) to (c6) and (d1) to (d6) correspond to $\widetilde{U}_1, \ldots, \widetilde{U}_{300}$, (e1) to (e6) and (f1) to (f6) correspond to $\widehat{U}_1, \cdots, \widehat{U}_{300}$, and (g1) to (g6) correspond to $U_1^\dagger, \ldots, U_{300}^\dagger$.

We make several observations based on Figure 1: First observe that as plots (b5), (c5), (d5), (e5), (f5), and (g5) show, when $n = 500$ and $h = H1$, all methods perform much better than the asymptotic theory. However, plots (e1) and (f1), as well as those in (e3) and (f3), show that with smaller sample sizes ($n = 50$ and 100) and $h = H1$, the weighted bootstrap statistic (16) performs slightly better than the percentile bootstrap, the weighted bootstrap (15), and Dirichlet random weighting. This is also intuitively reasonable because, in practice, the statistic (16) (and not (15) or (12)) is the right bootstrap counterpart for the statistics in (5). Figure 1 also shows that the bandwidth $H1$ yields far better performance than $H2$. Figure 2 presents the same results when f is a Normal density with $\mu = 0.5$ and $\sigma = 0.2$. Again, we observe that the weighted bootstrap has the ability to perform quite well.

Figures 1 and 2 only provide an informal way of assessing various methods. Therefore, in addition to these graphical methods, we have also carried out formal goodness-of-fit tests of hypotheses to

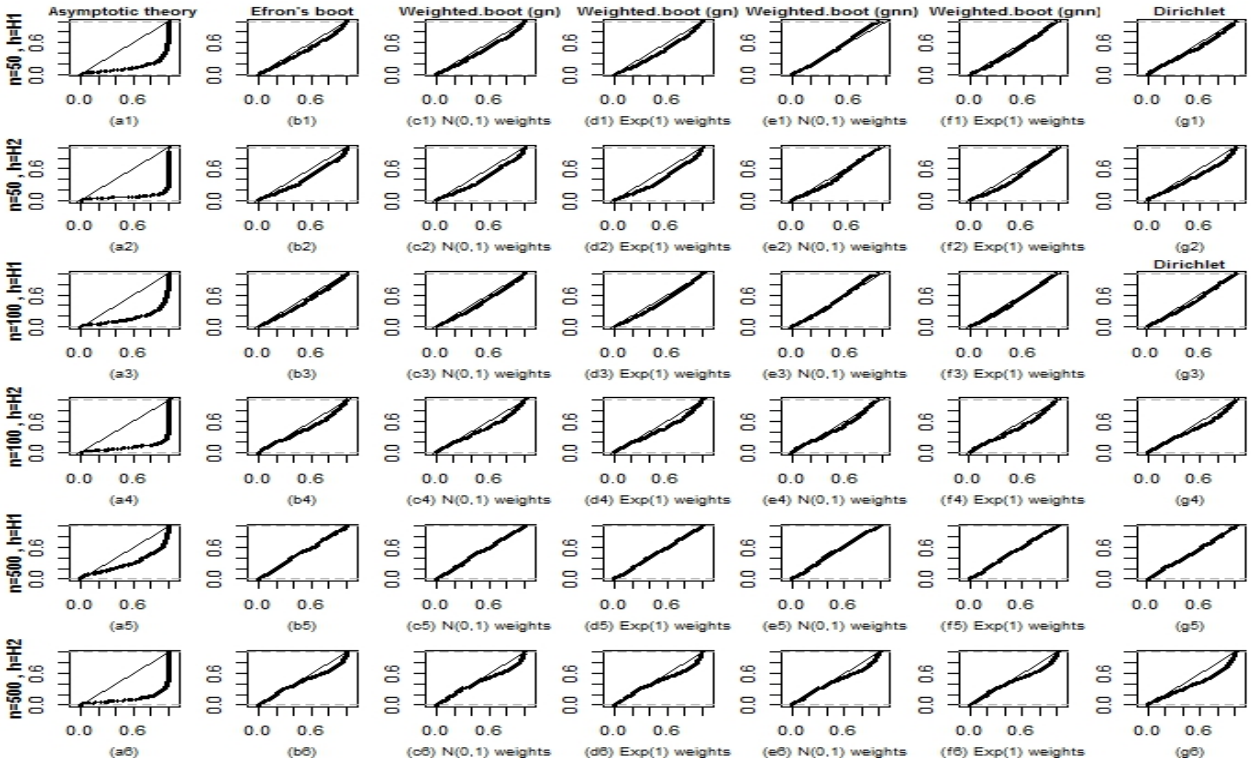

Figure 2: Plots of empirical cdf's when f is the density of a Normal distribution with $\mu = 0.5$ and $\sigma = 0.2$. Here, plots (a1) to (a6) correspond to $U_1, \cdots, U_{300}$, (b1) to (b6) correspond to $U_1^*, \cdots, U_{300}^*$, (c1) to (c6) and (d1) to (d6) correspond to $\widetilde{U}_1, \ldots, \widetilde{U}_{300}$, (e1) to (e6) and (f1) to (f6) correspond to $\widehat{U}_1, \cdots, \widehat{U}_{300}$, and (g1) to (g6) correspond to $U_1^\dagger, \ldots, U_{300}^\dagger$.

assess the performance of all the methods discussed above. More specifically, we carried out the classical Kolmogorov-Smirnov tests of:

$H_0^{(A)}$: $U_1, \ldots, U_{300}$ are iid Unif $[0, 1]$
$H_0^{(B)}$: $U_1^*, \ldots, U_{300}^*$ are iid Unif $[0, 1]$
$H_0^{(C)}$: $\widetilde{U}_1, \ldots, \widetilde{U}_{300}$ are iid Unif $[0, 1]$
$H_0^{(D)}$: $\widehat{U}_1, \ldots, \widehat{U}_{300}$ are iid Unif $[0, 1]$
$H_0^{(E)}$: $U_1^\dagger, \ldots, U_{300}^\dagger$ are iid Unif $[0, 1]$

under various setups. The results (p-values) are summarized in Table 1.

Table 1: P-values of Kolmogorov-Smirnov tests of uniformity in $H_0^{(A)}$, $H_0^{(B)}$, $H_0^{(C)}$, $H_0^{(D)}$, and $H_0^{(E)}$ under different setup. P-values that are larger than 10% appear in boldface. A larger p-value indicates that the corresponding method works better under the particular values of n and $h = H1$ or $H2$. The top half of the table corresponds to the case where f is the pdf of a Gamma distribution with shape parameter 5 and scale parameter 0.1, and the bottom half corresponds to f being the pdf of a normal distribution with $\mu = 0.5$ and $\sigma = 0.2$.

	$n = 50$		$n = 100$		$n = 500$	
	$H1$	$H2$	$H1$	$H2$	$H1$	$H2$
Method:						
Asymptotic theory via (5)	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}
Percentile bootstrap via (12)	1×10^{-4}	2×10^{-11}	0.0033	2×10^{-12}	**0.2591**	2×10^{-16}
Weighted bootstrap via (15) with N$(0,1)$ weights	3×10^{-5}	4×10^{-11}	0.0025	4×10^{-12}	**0.3070**	2×10^{-16}
Weighted bootstrap via (15) with Exp(1) weights	2×10^{-5}	5×10^{-13}	2×10^{-4}	3×10^{-13}	**0.3243**	2×10^{-16}
Weighted bootstrap via (16) with N$(0,1)$ weights	**0.1270**	8×10^{-5}	**0.2492**	2×10^{-6}	**0.6020**	2×10^{-15}
Weighted bootstrap via (16) with Exp(1) weights	0.0210	1×10^{-7}	0.0277	9×10^{-9}	**0.3800**	2×10^{-16}
Dirichlet random weight	7×10^{-5}	2×10^{-5}	0.0092	6×10^{-11}	5×10^{-4}	2×10^{-16}
Asymptotic theory via (5)	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}	2×10^{-16}
Percentile bootstrap via (12)	3×10^{-4}	9×10^{-7}	0.0310	9×10^{-7}	**0.8688**	6×10^{-8}
Weighted bootstrap via (15) with N$(0,1)$ weights	2×10^{-4}	5×10^{-7}	0.0277	2×10^{-6}	**0.7853**	6×10^{-8}
Weighted bootstrap via (15) with Exp(1) weights	2×10^{-5}	3×10^{-8}	0.0084	6×10^{-7}	**0.8928**	2×10^{-7}
Weighted bootstrap via (16) with N$(0,1)$ weights	0.0217	0.0033	**0.1103**	0.0107	**0.4413**	3×10^{-6}
Weighted bootstrap via (16) with Exp(1) weights	0.0792	2×10^{-5}	**0.1207**	1×10^{-4}	**0.7867**	9×10^{-7}
Dirichlet random weight	0.0083	3×10^{-6}	**0.1017**	1×10^{-5}	**0.6911**	4×10^{-9}

Here, the top half of Table 1 corresponds to the case where f is the pdf of a Gamma distribution with shape parameter 5 and scale parameter 0.1; in the bottom half of the table f is the pdf of a normal distribution with $\mu = 0.5$ and $\sigma = 0.2$. Our results show that the p-values corresponding to $H_0^{(A)}$ are all virtually zero under every setup, confirming the rather poor performance of the asymptotic result (9) in Theorem 1. Similarly, all test results are significant (small p-values) when $h = H2$. On the other hand, the boldfaced numbers in the second last column of Table 1 show that when $h = H1$ and $n = 500$, most of the p-values are quite large (> 0.25), confirming the good performance of various methods used; this is also confirmed by plots (b5), (c5), (d5), (ef), (f5) in both Figures 1 and 2 as well as plot (g5) in Figure 2. The situation, however, is not quite the same for smaller sample sizes. In fact, as Table 1 shows, when $n =$ 50 or $n =$ 100, all

p-values are small except for the five boldfaced values 0.1270, 0.2492, 0.1103, 0.1207, and 0.1017, indicating the better performance of the weighted bootstrap (16) and, to some extent, that of Dirichlet random weights (the last line of Table 1).

Next, we used our results to construct 90% confidence bands for the true density f (based on 300 Monte Carlo runs) under different settings. This resulted in 300 bands for each setup. Table 2 summarizes the coverage probabilities for each method under different values of n. Here, coverage is measured as the proportion of bands (out of 300 bands) that actually captured the true density f in the interval $[0, 1]$. Table 2 also gives the average areas, over 300 runs, of confidence bands constructed under each setup; these average areas appear in brackets. The top half of the table corresponds to the case where f is the pdf of a Gamma distribution with shape parameter 5 and scale parameter 0.1, and the bottom half corresponds to the case where f is the pdf of a Normal distribution with $\mu = 0.5$ and $\sigma = 0.2$.

Table 2: The following table presents the actual coverages of various confidence bands (measured as the proportion of 300 bands that captured the true pdf). The numbers appearing in brackets are the average areas of the confidence bands (averaged over 300 bands). The boldfaced values are those actual coverages that are closest to the 90% nominal coverage probability (they fall within a $[90 \pm 1]$% range). Here, the top half of the table corresponds to the case where f is the pdf of a Gamma distribution with shape parameter 5 and scale parameter 0.1, and the bottom half corresponds to the case where f is the pdf of a Normal distribution with $\mu = 0.5$ and $\sigma = 0.2$.

	$n = 50$		$n = 100$		$n = 500$	
	$H1$	$H2$	$H1$	$H2$	$H1$	$H2$
Method						
Asymptotic theory via (5)	0.267 (1.171)	0.153 (0.538)	0.323 (1.067)	0.180 (0.482)	0.500 (0.882)	0.167 (0.366)
Percentile bootstrap via (12)	0.802 (2.412)	0.727 (1.536)	0.857 (1.914)	0.723 (1.196)	0.880 (1.303)	0.653 (0.726)
Weighted bootstrap via (15) with N$(0,1)$ weights	0.807 (2.411)	0.730 (1.540)	0.853 (1.915)	0.717 (1.199)	0.880 (1.304)	0.650 (0.726)
Weighted bootstrap via (15) with Exp(1) weights	0.817 (4.453)	0.720 (2.624)	0.860 (1.943)	0.723 (1.210)	0.883 (1.306)	0.667 (0.727)
Weighted bootstrap via (16) with N$(0,1)$ weights	0.946 (2.449)	0.930 (1.549)	0.947 (2.464)	0.863 (1.511)	**0.897** (1.387)	0.737 (0.768)
Weighted bootstrap via (16) with Exp(1) weights	0.887 (2.958)	0.843 (1.842)	**0.897** (2.119)	0.783 (1.321)	**0.893** (1.344)	0.680 (0.748)
Dirichlet random weight	0.798 (3.035)	0.783 (1.533)	0.876 (2.093)	0.787 (1.262)	**0.890** (1.331)	0.697 (0.713)
Asymptotic theory via (5)	0.273 (1.353)	0.121 (0.435)	0.347 (1.282)	0.187 (0.373)	0.563 (1.046)	0.250 (0.233)
Percentile bootstrap via (12)	0.812 (2.719)	0.770 (1.448)	0.859 (2.194)	0.803 (1.086)	**0.897** (1.487)	0.747 (0.575)
Weighted bootstrap via (15) with N$(0,1)$ weights	0.813 (2.716)	0.767 (1.449)	0.857 (2.192)	0.810 (1.087)	**0.890** (1.486)	0.733 (0.575)
Weighted bootstrap via (15) with Exp(1) weights	0.813 (4.865)	0.773 (2.672)	0.863 (2.220)	0.810 (1.092)	**0.900** (1.493)	0.740 (0.577)
Weighted bootstrap via (16) with N$(0,1)$ weights	0.983 (2.767)	0.937 (1.461)	0.963 (2.953)	0.950 (1.432)	**0.910** (1.595)	0.780 (0.607)
Weighted bootstrap via (16) with Exp(1) weights	0.917 (3.359)	0.867 (1.755)	**0.910** (2.453)	0.873 (1.207)	**0.906** (1.542)	0.757 (0.593)
Dirichlet random weight	0.914 (3.650)	0.873 (1.816)	**0.907** (2.389)	0.887 (1.209)	**0.903** (1.541)	0.763 (0.602)

The boldfaced values in the table identify those actual coverages that are very close to the 90% nominal coverage; here an actual coverage is considered to be close to the nominal 90% level if it falls within the [90$\pm$1]% range. Table 2 shows that the result are substantially better when $H1$ is used for the bandwidth; this is also consistent with our findings in Table 1, Figure 1, and Figure 2, where all approximations work better when $h = H1$ is chosen.

The results in Table 2, Table 1, Figure 1, and Figure 2 also show that the weighted bootstrap has the ability to perform quite well; this is particularly true when the "studentized" type weighted bootstrap statistic $\widehat{\Gamma}_{nn}(t, \ell)$, as defined by (16), is used. This is not a fluke. In fact, as mentioned

in Remark 1, the statistic $\widehat{\Gamma}_{nn}(t,\ell)$, (and not $\widetilde{\Gamma}_{nn}(t,\ell)$ in (15)), is the correct weighted bootstrap counterpart of $\widehat{\Gamma}_n(t,\ell)$. Of course, confidence bands based on (16) are typically "wider" (have larger areas) than those based on (15), which is similar to using t instead of Z values in a confidence interval for a population mean μ.

Remark 2 *[The number of Monte Carlo runs.]*
We have carried out our numerical examples based on $N = 300$ Monte Carlo runs. Our choice of N is based on the variance of the Monte Carlo estimator. More specifically, for $j = 1, \ldots, N$, define

$$Y_j = \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ confidence band captures the true density } f \text{ uniformly in } [0,1], \\ 0 & \text{otherwise.} \end{cases}$$

Here $Y_1, \ldots, Y_N$ are iid Bernoulli(p) random variables, generated by the N simulated data sets of size n each, where $p = E(Y_j)$ can be estimated by $\widehat{p} = \sum_{j=1}^{N} Y_j/N$. In fact, this is how the coverage probabilities in Table 2 are computed. But $\mathrm{Var}(\widehat{p}) = p(1-p)/N \leq \frac{1}{4N}$, and if this variance is required to be less than a pre-specified threshold δ, then we can achieve this by choosing $N \geq \frac{1}{4\delta}$. Here, we took $\delta = 10^{-3}$ which leads to requiring $N \geq 250$. Therefore, our choice of $N = 300$ is more than sufficient. We must also add that our initial small pilot study based on several values of N shows that once N reaches about 200, the main message of tables 1 and 2 will not change (by increasing N) in the sense that the positions of the boldfaced values will remain the same in both tables. Of course, a much larger N can render more stable (less variable) results, but it comes at a formidable computational cost. Also, in our view, a sample of size $N = 300$ is large enough to warrant the applications of the classical Kolmogorove-Smirnov test.

3 Proofs

PROOF OF THEOREM 2. Part (i).
Let $G(x) = P\{X \leq x\}$ and $G_n(x) = n^{-1}\sum_{i=1}^{n} I\{X_i \leq x\}$. Also put

$$\begin{aligned} G_{nn}(x) &= n^{-1}\sum_{i=1}^{n}(1+\xi_i-\bar{\xi})I\{X_i \leq x\} \\ \beta_n(x) &= \sqrt{n}\,(G_{nn}(x) - G_n(x)) = n^{-1/2}\sum_{i=1}^{n}(\xi_i-\bar{\xi})I\{X_i \leq x\} \end{aligned}$$

and observe that with $L_n^{(\ell)}(\cdot)$ as in (3), we have

$$\begin{aligned} f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t) &= \frac{1}{nh^{\ell+1}}\left[\sum_{i=1}^{n}(1+\xi_i-\bar{\xi})\,L_n^{(\ell)}\left(\frac{t-X_i}{h}\right) - \sum_{i=1}^{n} L_n^{(\ell)}\left(\frac{t-X_i}{h}\right)\right] \\ &= \frac{1}{n^{1/2}h^{\ell+1}}\int_{\mathbb{R}} L_n^{(\ell)}\left(\frac{t-u}{h}\right)\,d\beta_n(u)\,. \end{aligned}$$

Consequently,

$$n^{1/2}h^{\beta+\ell+1/2}\left(f_{nn}^{(\ell)}(t)-f_n^{(\ell)}(t)\right)=h^{\beta-1/2}\int_{\mathbb{R}}L_n^{(\ell)}\left(\frac{t-u}{h}\right)\,d\beta_n(u)\,.$$

Next, we state a result of Horváth et al. (2000) on the best approximation of the process $\{\beta_n(t),\ t\in\mathbb{R}\}$ by a sequence of Brownian bridges. A powerful generalization of this

Lemma 1 *[Horváth et al. (2000; Theorem 1.3).] Let $\beta_n(x)=n^{-1/2}\sum_{i=1}^n(\xi_i-\bar{\xi})I\{X_i\le x\}$ and suppose that the weights $\xi_1,\ldots,\xi_n$ satisfy assumption A4. Then there exists a sequence of Brownian bridges $\{B_n(t),\ 0\le t\le 1\}$ such that*

$$P\left\{\sup_{-\infty<x<\infty}|\beta_n(x)-B_n(G(x))|>n^{-1/2}(c_1\log n+y)\right\}\le c_2e^{-c_3y},$$

for all $y\ge 0$, where c_1, c_2 and c_3 are positive constants.

An application of the above lemma together with the Borel-Cantelli lemma immediately gives

$$\sup_{-\infty<x<\infty}|\beta_n(x)-B_n(G(x))|=_{\text{a.s.}}\mathcal{O}(n^{-1/2}\log n). \tag{19}$$

Now observe that with $B_n(\cdot)$ as above,

$$\begin{aligned}
&\sup_{0\le t\le 1}\left|\frac{n^{1/2}h^{\beta+\ell+1/2}}{\sqrt{g(t)}}\left(f_{nn}^{(\ell)}(t)-f_n^{(\ell)}(t)\right)-\frac{h^{\beta-1/2}}{\sqrt{g(t)}}\int_{\mathbb{R}}L_n^{(\ell)}\left(\frac{t-u}{h}\right)\,dB_n(G(u))\right|\\
&=\sup_{0\le t\le 1}\left|\frac{h^{\beta-1/2}}{\sqrt{g(t)}}\left[\int_{\mathbb{R}}L_n^{(\ell)}\left(\frac{t-u}{h}\right)\,d\beta_n(u)-\int_{\mathbb{R}}L_n^{(\ell)}\left(\frac{t-u}{h}\right)\,dB_n(G(u))\right]\right|\\
&=\sup_{0\le t\le 1}\left|\frac{h^{\beta-1/2}}{\sqrt{g(t)}}\int_{\mathbb{R}}\Big(\beta_n(t-hs)-B_n(G(t-hs))\Big)\,L_n^{(\ell+1)}(s)\,ds\right|\\
&\le h^{-1/2}\sup_{0\le x\le 1}\big|\beta_n(x)-B_n(G(x))\big|\times\sup_{0\le t\le 1}\frac{h^{\beta}}{\sqrt{g(t)}}\int_{\mathbb{R}}\left|L_n^{(\ell+1)}(s)\right|ds\\
&=_{\text{a.s.}}\mathcal{O}\left(\log n/\sqrt{nh}\right)\,,
\end{aligned} \tag{20}$$

where the last line follows from (19), part (a) of assumption A2, and part (a) of assumption A3. Bissantz et al. (2007) studied the asymptotic behavior of the process

$$\Gamma_{n,0}(t,\ell)\ :=\ \frac{h^{\beta-1/2}}{\sqrt{g(t)}}\int_{\mathbb{R}}L_n^{(\ell)}\left(\frac{t-u}{h}\right)\,dB_n(G(u))\,,\quad t\in[0,1]\,.$$

Their results show that $(2\log(1/h))^{1/2}\{C_{1,\ell}^{-1/2}\sup_{0\le t\le 1}\left|\Gamma_{n,0}(t,\ell)\right|-d_{n,\ell}\}\to_d Y$, where $P\{Y\le y\}=\exp(-2e^{-y})$. Putting this together with (20), we obtain

$$\sqrt{2\log(1/h)}\left(C_{1,\ell}^{-1/2}\sup_{0\le t\le 1}\left|\frac{n^{1/2}h^{\beta+\ell+1/2}}{\sqrt{g(t)}}\left(f_{nn}^{(\ell)}(t)-f_n^{(\ell)}(t)\right)\right|-d_{n,\ell}\right)\longrightarrow_d Y\,, \tag{21}$$

as $n \to \infty$, where $P\{Y \le y\} = \exp(-2e^{-y})$. To complete the proof of part (i) of Theorem 2 we need to show that

$$\sqrt{\log(1/h)}\, n^{1/2} h^{\beta+\ell+1/2} \sup_{0\le t\le 1} \left| \left[\frac{1}{\sqrt{\widehat{g}_n(t)}} - \frac{1}{\sqrt{g(t)}} \right] \left(f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t) \right) \right| \longrightarrow_p 0\,, \tag{22}$$

as $n \to \infty$. To show this, first observe that the expression on the left side of the arrow in (22) is bounded by

$$\sqrt{\log(1/h)}\, n^{1/2} h^{\beta+\ell+1/2} \sup_{0\le t\le 1} \left| \left(f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t) \right) / \sqrt{g(t)} \right| \times \sup_{0\le t\le 1} \left| \left(\widehat{g}_n(t) - g(t) \right) / \widehat{g}_n(t) \right| . \tag{23}$$

However, in view of (21), we have

$$\sqrt{\log(1/h)}\, n^{1/2} h^{\beta+\ell+1/2} \sup_{0\le t\le 1} \left| \left(f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t) \right) / \sqrt{g(t)} \right| = \mathcal{O}_p(\log(1/h))\,. \tag{24}$$

Furthermore, (6) together with part (a) of assumption A3 imply (see the Appendix) that

$$\sup_{0\le t\le 1} \left| \left(\widehat{g}_n(t) - g(t) \right) / \widehat{g}_n(t) \right| = o_p\big(\big(\log(1/h) \big)^{-1} \big)\,. \tag{25}$$

Now, (22) follows from (23), (24), and (25), which completes the proof of Part (i).

Part (ii).
Let $\widetilde{\Gamma}_{nn}(t,\ell)$ and $\widehat{\Gamma}_{nn}(t,\ell)$ be as in (15) and (16). Then, in view of Part (i) and the fact that $\left| \widehat{\Gamma}_{nn}(t,\ell) - \widetilde{\Gamma}_{nn}(t,\ell) \right| = n^{1/2} h^{\beta+\ell+1/2} \left| \left(f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t) \right) \times \left\{ \widehat{g}_{nn}^{-1/2}(t) - \widehat{g}_n^{-1/2}(t) \right\} \right|$, where $\widehat{g}_n$ and $\widehat{g}_{nn}$ are as in (17) and (18) respectively, it is sufficient to show that

$$\sqrt{\log(1/h)}\, n^{1/2} h^{\beta+\ell+1/2} \sup_{0\le t\le 1} \left| \frac{f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t)}{\sqrt{\widehat{g}_n(t)}} \right| \cdot \sup_{0\le t\le 1} \left| \frac{\widehat{g}_{nn}(t) - \widehat{g}_n(t)}{\widehat{g}_{nn}(t)} \right| \longrightarrow_p 0\,, \tag{26}$$

as $n \to \infty$. But, by Part (i) we obtain

$$\sqrt{\log(1/h)}\, n^{1/2} h^{\beta+\ell+1/2} \sup_{0\le t\le 1} \left| \frac{f_{nn}^{(\ell)}(t) - f_n^{(\ell)}(t)}{\sqrt{\widehat{g}_n(t)}} \right| \overset{(15)}{=} \sqrt{\log(1/h)} \sup_{0\le t\le 1} \left| \widetilde{\Gamma}_{nn}(t,\ell) \right| = \mathcal{O}_p(\log(1/h))\,. \tag{27}$$

To deal with the second supremum term in (26), let $B_n(\cdot)$ and $\beta_n(\cdot)$ be as in Lemma 1 and observe that since $(n\lambda_n)^{1/2}\big(\widehat{g}_{nn}(t) - \widehat{g}_n(t)\big) = \lambda_n^{-1/2} \int_{\mathbb{R}} \mathcal{H}((t-s)/\lambda_n)\, d\beta_n(s)$, part (a) of assumption A3 and arguments similar to (and, in fact, easier than) those that lead to (20) yield

$$\sup_{0\le t\le 1} \left| \sqrt{\frac{n\lambda_n}{g(t)}}\, \left(\widehat{g}_{nn}(t) - \widehat{g}_n(t) \right) - \frac{1}{\sqrt{\lambda_n g(t)}} \int_{\mathbb{R}} \mathcal{H}\left(\frac{t-s}{\lambda_n} \right) dB_n(G(s)) \right| =_{\text{a.s.}} \mathcal{O}\left(\frac{\log n}{\sqrt{n\lambda_n}} \right) \tag{28}$$

Now let $\{B(t),\ 0 \le t \le 1\}$ be a Brownian bridge and note that for each $n = 1, 2, \ldots$

$$\left\{ \frac{1}{\sqrt{\lambda_n g(t)}} \int_{\mathbb{R}} \mathcal{H}\left((t-s)/\lambda_n\right)\, dB_n(G(s)), \quad 0 \le t \le 1 \right\}$$
$$=_d \left\{ \frac{1}{\sqrt{\lambda_n g(t)}} \int_{\mathbb{R}} \mathcal{H}\left((t-s)/\lambda_n\right)\, dB(G(s)), \quad 0 \le t \le 1 \right\}$$

On the other hand, Bickel and Rosenblatt (1973) studied the behavior of the process $\Lambda_n(t) = (\lambda_n g(t))^{-1/2} \int_{\mathbb{R}} \mathcal{H}\left((t-s)/\lambda_n\right)\, dB(G(s))$, $0 \le t \le 1$, and showed that

$$P\left\{ \sqrt{2\delta \log n}\, \left(\eta^{-1/2} \sup_{0 \le t \le 1} |\Lambda_n(t)| - \varrho_n \right) \le z \right\} \longrightarrow \exp\left(-2e^{-z}\right), \tag{29}$$

where $\eta = \int \mathcal{H}^2(u)\, du$, and

$$\varrho_n \;=\; \sqrt{2\delta\, \log n} + \begin{cases} \frac{\log C_1 - \frac{1}{2}\, \log \pi + \frac{1}{2}(\log \delta + \log\log n)}{(2\delta\, \log n)^{1/2}}, & \text{if } C_1 := \frac{\mathcal{H}^2(A) + \mathcal{H}^2(-A)}{2\eta} > 0, \\ \frac{\log\left[(1/\pi)(C_2/2)^{1/2}\right]}{(2\delta\, \log n)^{1/2}}, & \text{otherwise}, \end{cases}$$

and where $C_2 = \frac{1}{2\eta} \int [\mathcal{H}'(t)]^2 dt$. Therefore, in view of (28),

$$P\left\{ \sqrt{2\delta \log n}\, \left(\eta^{-1/2} \sup_{0 \le t \le 1} \left| \sqrt{n\lambda_n / g(t)}\, \left(\widehat{g}_{nn}(t) - \widehat{g}_n(t)\right) \right| - \varrho_n \right) \le z \right\} \longrightarrow \exp\left(-2e^{-z}\right),$$

which, together with part (a) of assumption A3, yield

$$\sup_{0 \le t \le 1} |\widehat{g}_{nn}(t) - \widehat{g}_n(t)| = \mathcal{O}_p\left(\sqrt{\log n/(n\lambda_n)} \right). \tag{30}$$

Putting together (30) and part (a) of assumption A3, it can be shown (see the Appendix) that

$$\sup_{0 \le t \le 1} \left| \left[\widehat{g}_{nn}(t) - \widehat{g}_n(t)\right] / \widehat{g}_{nn}(t) \right| = \mathcal{O}_p\left(\sqrt{\log n/(n\lambda_n)} \right). \tag{31}$$

This together with (27), and the condition that $\log(1/h)\, \sqrt{\log n/(n\lambda_n)} \to 0$, imply that

$$\big(\text{left side of the arrow in (26)}\big) \;=\; \mathcal{O}_p\left(\log(1/h)\right) \cdot \mathcal{O}_p\left(\sqrt{\log n/(n\lambda_n)} \right) \;=\; o_p(1),$$

as $n \to \infty$. This completes the proof of Part (ii) of Theorem 2.

□

Appendix

DERIVATION OF (25):
The derivation of (25) is similar to, and easier than, that of (31) given below.

Since $\left|\widehat{g}_n(t) - g(t)\right| \geq \left(\left|\widehat{g}_n(t) - g(t)\right| / |\widehat{g}_n(t)|\right) \cdot \inf_{t\in[0,1]} |\widehat{g}_n(t)|$, we find that for all $n > 0$

$$\sup_{t\in[0,1]} \left|\widehat{g}_n(t) - g(t)\right| \;\geq\; \sup_{t\in[0,1]} \left|\frac{\widehat{g}_n(t) - g(t)}{\widehat{g}_n(t)}\right| \cdot \inf_{t\in[0,1]} \left|\widehat{g}_n(t)\right|. \tag{32}$$

Next, observe that for all $n > 0$

$$-\sup_{t\in[0,1]} \left|\widehat{g}_n(t) - g(t)\right| \leq \inf_{t\in[0,1]} \left(|\widehat{g}_n(t)| - g(t)\right) \leq \inf_{t\in[0,1]} |\widehat{g}_n(t) - g(t)| \leq \sup_{t\in[0,1]} \left|\widehat{g}_n(t) - g(t)\right|.$$

But, as $n \to \infty$, both the far left and the far right sides of the above chain of inequalities converge to zero (by (6)); hence, so does the middle term, i.e., $\inf_{t\in[0,1]} \left(|\widehat{g}_n(t)| - g(t)\right) \to_p 0$. Consequently, in view of part (a) of assumption A3, we find

$$\lim_{n\to\infty} \inf_{t\in[0,1]} |\widehat{g}_n(t)| \geq \lim_{n\to\infty} \left\{ \inf_{t\in[0,1]} \left(|\widehat{g}_n(t)| - g(t)\right) + \inf_{t\in[0,1]} g(t) \right\} =_p 0 + \inf_{t\in[0,1]} g(t) > 0.$$

Now, multiplying both sides of the inequality in (32) by $\log(1/h)$ and taking the limit as $n \to \infty$, we find (via (6)) that $\log(1/h) \sup_{t\in[0,1]} \left|\left(\widehat{g}_n(t) - g(t)\right) / \widehat{g}_n(t)\right| \to_p 0$, which is the same as (25). □

DERIVATION OF (31):
Since $\left|\widehat{g}_{nn}(t) - \widehat{g}_n(t)\right| \geq \left(\left|\widehat{g}_{nn}(t) - \widehat{g}_n(t)\right| / \left|\widehat{g}_{nn}(t)\right|\right) \cdot \inf_{t\in[0,1]} |\widehat{g}_{nn}(t)|$, we find that for all $n > 0$

$$\sup_{t\in[0,1]} \left|\widehat{g}_{nn}(t) - g_n(t)\right| \;\geq\; \sup_{t\in[0,1]} \left|\frac{\widehat{g}_{nn}(t) - g_n(t)}{\widehat{g}_{nn}(t)}\right| \cdot \inf_{t\in[0,1]} |\widehat{g}_{nn}(t)| \tag{33}$$

Now, observe that for all $n > 0$

$$-\sup_{t\in[0,1]} \left|\widehat{g}_{nn}(t) - \widehat{g}_n(t)\right| \leq \inf_{t\in[0,1]} \left(|\widehat{g}_{nn}(t)| - |\widehat{g}_n(t)|\right) \leq \inf_{t\in[0,1]} \left|\widehat{g}_{nn}(t) - \widehat{g}_n(t)\right| \leq \sup_{t\in[0,1]} \left|\widehat{g}_{nn}(t) - \widehat{g}_n(t)\right|.$$

But, as $n \to \infty$, both the far left and the far right sides of the above chain of inequalities converge to zero (by (30)) and, hence, so does the middle term, i.e., $\inf_{t\in[0,1]} \left(|\widehat{g}_{nn}(t)| - |\widehat{g}_n(t)|\right) \to_p 0$. This together with part (a) of assumption A3 and the fact that $\sup_{t\in[0,1]} |\widehat{g}_n(t) - g(t)| \to_p 0$ for the kernel density estimator $\widehat{g}_n$ defined in (17), imply that

$$\begin{aligned}
\lim_{n\to\infty} \inf_{t\in[0,1]} |\widehat{g}_{nn}(t)| &= \lim_{n\to\infty} \inf_{t\in[0,1]} \left\{|\widehat{g}_{nn}(t)| - |\widehat{g}_n(t)| + |\widehat{g}_n(t)| - g(t) + g(t)\right\} \\
&\geq \lim_{n\to\infty} \left\{ \inf_{t\in[0,1]} \left(|\widehat{g}_{nn}(t)| - |\widehat{g}_n(t)|\right) + \inf_{t\in[0,1]} \left(|\widehat{g}_n(t)| - g(t)\right) + \inf_{t\in[0,1]} g(t) \right\} \\
&=_p 0 + 0 + \inf_{t\in[0,1]} g(t) \;>\; 0\,,
\end{aligned}$$

where we have used the simple inequality that for bounded functions $\psi_1, \psi_2 : \mathcal{C} \to \mathbb{R}$, where $\mathcal{C} \subset \mathbb{R}$, one has $\inf_{t\in\mathcal{C}}\{\psi_1(t) + \psi_2(t)\} \geq \inf_{t\in\mathcal{C}} \psi_1(t) + \inf_{t\in\mathcal{C}} \psi_2(t)$. Next let $\{a_n\}$ be any sequence of

positive numbers converging to zero and observe that if we multiply both sides of the inequality (33) by $a_n\sqrt{n\lambda_n/\log n}$, then upon taking the limit as $n \to \infty$, we find (in view of (30)) that $a_n\sqrt{n\lambda_n/\log n}\,\sup_{t\in[0,1]}\left|\left(\widehat{g}_{nn}(t)-\widehat{g}_n(t)\right)/\,\widehat{g}_{nn}(t)\right| \to_p 0$. Now, (31) follows because the sequence $\{a_n\}$ can be chosen to converge arbitrarily slowly.

□

References

Achilleos A, Delaigle A (2012) Local bandwidth selectors for deconvolution kernel density estimation. Stat and Computing 22:563-577

Ahlgren N, Catani P (2016) Wild bootstrap tests for autocorrelation in vector autoregressive models. Stat Papers 58:1189-1216

Barbe P. and Bertail P (1995) The Weighted Bootstrap. Lecture Notes in Statistics, Springer

Benšić M, Sabo (2016) K Uniform distribution width estimation from data observed with Laplace additive error. J Korean Stat Society 45:505-517

Bickel P, Rosenblatt M (1973) On some global measures of the deviations of density function estimates. Ann of Stat 1:1075-1095

Bissantz N, Dümbgen, L, Holzmann, H, and Munk, A (2007) Nonparametric confidence bands in deconvolution density estimation. J. R. Stat. Soc. Ser. B. 69:483-506

Bouzebda S, Papamichail C, Limnios N (2017) On a multidimensional general bootstrap for empirical estimator of continuous-time semi-Markov kernels with applications. J Nonparametric Stat. https://doi.org/10.1080/10485252.2017.1404059.

Burch C, Guyader S, Samarov D, and Shen H (2007) Experimental estimate of the abundance and effects of nearly neutral mutations in the RNA virus ϕ6. Genetics 176:467-476

Burke M D (2010) Approximations for a multivariate hybrid process with applications to change-point detection. Mathematical Methods of Statistics, 19:121-135.

Burke M (1998). 'A Gaussian bootstrap approach to estimation and tests' in Asymptotic Methods in Probability and Statistics, ed. B. Szyszkowicz, North-Holland, Amsterdam pp. 697-706

Burke M (2000). Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. Stat and Probab Lett, 46:13-20

Carroll R, Hall P (1988) Optimal rates of convergence for deconvolving a density. J Am Stat Assoc 83:1184-1186

Cator E (2001) Deconvolution with arbitrary smooth kernels. Stat Probab Lett 54:205-215

Cheng G, Huang JZ (2010) Bootstrap consistency for general semiparametric M-estimation. Ann of Stat 38:2884-2915

Chesneau C, Fadili J (2013) Wavelet-based density estimation in a heteroscedastic convolution model. Comm Stat Theory & Method 42:3085-3099

Chiang CT, Wang SH, Hung H (2009) Random weighting and Edgeworth expansion for the non-parametric time-dependent AUC estimator. Stat Sinica 19: 969-979

Chiang CT, James LF, Wang MC (2005) Random weighted bootstrap method for recurrent events with informative censoring. Lifetime Data Anal 11:489-509

Cornelis G, Hassellöv M (2014) A signal deconvolution method to discriminate smaller nanoparticles in single particle ICP-MS. J. Analytical Atomic Spectrometry 29:134-144

Delaigle A, Gijbels I (2004) Practical bandwidth selection in deconvolution kernel density estimation. Comput Stat Data Anal 45:249-267

Delaigle A, Gijbels I (2006) Data-driven boundary estimation in deconvolution problems. Comput Stat & Data Analysis 50:1965-1994

Delaigle A, Hall P (2014) Parametrically assisted nonparametric estimation of a density in the deconvolution problem. J American Stat Assoc 109:717-729

Delaigle A, Meister A (2008) Density estimation with heteroscedastic error. Bernoulli 14: 562-579

Efron B (1979) Bootstrap methods: another look at the jackknife. Ann of Stat 7:1-26

van Es B, Uh H (2004) Asymptotic normality of nonparametric kernel type deconvolution density estimators: Crossing the Cauchy boundary. J Nonparametric Stat 16:261-277

Van Es B, Uh H. (2005) Asymptotic normality of kernel-type deconvolution estimators. Scand J Stat 32:467-483

Fan J (1991) Asymptotic normality for deconvolution kernel density estimators. Sankhya Indian J Stat A 53:97-110

Fan J (1992). Deconvolution with Supersmooth Distributions. Canadian J of Stat 20:155-169

Fan Y, Liu Y (19987) A note on asymptotic normality for deconvolution kernel density estimators. Sankhya Indian J. Stat. A 59:138-141

Galvao A, Montes-Rojas G, Olmo J, Song S (2018) On solving endogeneity with invalid instruments: an application to investment equations, J. R. Statist. Soc. A 181:1-28

Gao S, Zhong Y (2010) The random weighting estimation of kernel density. J of Stat Plan Inference 140:2403-2407

Hall P (1992) The Bootstrap and Edgeworth Expansion. Springer-verlag, New York

Hall P (1991) On convergence rates of suprema. Prob Theory and Related Fields 89:447-455

Hall P, Mammen E. (1994) On general resampling algorithms and their performance in distribution

estimation, Ann of Stat, 22:2011-2030

Hesse CH (1999) Data-driven deconvolution. J Nonparametr Stat 10:343-373

Horváth L (2000). Approximations for hybrids of empirical and partial sums processes. J of Stat Plan and Infer, 88:1-18

Horváth L, Kokoszka P, Steinebach J (2000) Approximations for weighted bootstrap processes with an application. Stat and Probab Lett 48:59-70

Janssen A (2005) Resampling Student's t-type statistics. Ann of Inst Stat Math 57:507-529

Janssen A, Pauls T (2003) How do bootstrap and permutation tests work? Ann of Stat, 31:768-806

Kojadinovic I, Yan J (2012) Goodness-of-fit testing based on a weighted bootstrap: A fast large-sample alternative to the parametric bootstrap, Canadian J of Stat 40:480-500

Kojadinovic I, Yan J, Holmes M (2011) Fast large-sample goodness-of-fit for copulas, Statistica Sinica 21:841-871

Konakov V, Piterbarg V (1984) On the convergence rate of maximal deviation distribution for kernel regression estimates. J Multivariate Analysis 15:279-294

Kosorok M (2008) Introduction to Empirical Processes and Semiparametric Inference. Springer Series in Statistics, Springer-Verlag, New York

Kulik R (2008) Nonparametric deconvolution problem for dependent sequences. Electron J Stat 2:722-740

Lacour C (2006) Rates of convergence for nonparametric deconvolution. C. R. Acad. Sci. Paris Ser. I 342:877-882

Lee M, Shen H, Burch C, and Marron S (2010) Direct deconvolution density estimation of a mixture distribution motivated by mutation effects distribution. J Nonparametric Stat 22:1-22

Mason DM, Newton MA (1992) A rank statistics approach to the consistency of a general bootstrap. Ann of Stat 20:1611-1624

Masry E (1991) Multivariate probability density deconvolution for stationary random processes. IEEE Trans Inf Theory 37:1105-1115

McIntyre J, Stefanski L (2011) Density estimation with replicate heteroscedastic measurements. Ann Inst Stat Math 63: 81-99

Meister A (2010) Density deconvolution in a two-level heteroscedastic model with unknown error density. Electr J Stat 4:36-57

Meister A (2009) Deconvolution Problems in Nonparametric Statistics. Lecture notes in statistics, Vol. 193, Springer.

Mojirsheibani M, Pouliot W (2017) Weighted bootstrapped kernel density estimators in two-sample problems. J Nonparametric Stat 29:61-84

Praestgaard J, Wellner J A (1993) Exchangeably weighted bootstraps of the general empirical process. Ann of Probab 21:2053-2086

Rubin DB (1981) The Bayesian bootstrap. Ann. Statist., 9:130-134

Shao J, Tu DS (1995) The Jackknife and Bootstrap. Springer Series in Statistics, New York

Singh K (1981) On the asymptotic accuracy of Efron's bootstrap. Ann Stat 9:1187-1195.

Stefanski L (1990) Rates of convergence of some estimators in a class of deconvolution problems. Stat Probab Lett 9:229-235

Stefanski L, Carroll R (1990) Deconvoluting kernel density estimators. Statistics 21:169-184

Wand M (1998) Finite sample performance of deconvolving density estimators. Stat Probab Lett 37:131-139

Wang XF, Wang B (2011) Deconvolution estimation in measurement error models: The R package decon. J of Statistical Software 39:1-24

Weng CS (1989) On a second-order asymptotic property of the Bayesian bootstrap mean. Ann. Statist., 17:705-710.

Zamini R, Fakoor V, Sarmad M (2015) On estimation of a density function in multiplicative censoring. Stat Pap 56:661-676

Zhang CH (1990) Fourier methods for estimating mixing densities and distributions. Ann Stat 18:06-831

Zheng, ZG, Tu DS (1988). Random weighting methods in regression models. Sci. Sinica Ser. A, 31:1442-1459.