On a Weighted Bootstrap Approximation of the L_p Norms of Kernel Density Estimators

Bo Liu¹ and Majid Mojirsheibani^{2 3}

Department of Mathematics, California State University, Northridge, CA 91330

Abstract

A weighted bootstrap method is considered to approximate the distribution of the L_p norms of kernel density estimates. Here p is any number in $[1, \infty)$. Using a Komlós-Major-Tusnády type approximation (Komlós et al., 1975) for weighted bootstrap processes, due to Horváth et al. (2000), we establish unconditional bootstrap central limit theorems for these L_p statistics.

Keywords Kernel, Density, CLT, Brownian bridge, Bootstrap.

1 Introduction

Let $X_1, \dots, X_n$ be independently and identically distributed (i.i.d.) random variables with distribution function F and Lebesgue probability density function $f = F'$. The popular Parzen-Rosenblatt kernel estimator of f is given by (Parzen, 1962; Rosenblatt, 1956)

$$f_n(x) = (nh_n)^{-1} \sum_{i=1}^n K((x - X_i)/h_n),$$

where the kernel K is a function satisfying certain regularity conditions. These conditions depend on the type of theoretical properties that one intends to establish for f_n ; chapters 2 and 3 of the monograph by Prakasa Rao (1983) provide many of the classical results along these lines. Here, h_n is the smoothing parameter of the kernel and satisfies $h_n \rightarrow 0$, as $n \rightarrow \infty$. A measure of the overall performance of f_n is given by the L_p -norm type statistic

$$I_n(p) = \int_{\mathbb{R}} |f_n(t) - f(t)|^p d\mu(t), \quad 1 \leq p < \infty,$$

where μ is a measure on the Borel sets of $\mathbb{R}$. Bickel and Rosenblatt (1973) studied this statistic for the special case of $p = 2$ and gave a central limit theorem for $I_n(p)$ for the case of $d\mu(t) = a(t)dt$, where $a(t)$ is assumed to be a bounded piece-wise smooth integrable function. Much later, Csörgő and Horváth (1988) established central limit theorems for

¹Email: bo.liu.114@my.csun.edu.

² Corresponding author. Email: majid.mojirsheibani@csun.edu

³This work is supported in part by the NSF Grant DMS-1407400 of the second author.

$I_n(p)$, for all values of $p \in [1, \infty)$, without requiring μ to have a finite support. Horváth (1991) has extended these results to the multivariate case. Hall (1984) established central limit theorems for $I_n(p)$ in the special case of $p = 2$ when $\mu(t) = t$. The results of Csörgő and Horváth (1988) and Horváth (1991) are not restricted to the case of $\mu(t) = t$. In fact, the important case of $\mu(t) = F(t)$, i.e., when $I_n(p) = \int_{\mathbb{R}} |f_n(t) - f(t)|^p f(t) dt$, has been of particular interest in the literature; see, for example, (Wegman, 1972; Fryer, 1977; Steele, 1978; Hall, 1978). Other relevant results are those of Mojirsheibani (2007) who suggested methods for approximating the distribution of $I_n(p)$, as well as those of Berinet et al. (1995) who established central limit theorems for $I_n(1)$; also see (Beirlant and Mason, 1995).

From a goodness-of-fit point of view, the statistics $\int_{\mathbb{R}} |f_n(t) - f_0(t)|^p d\mu(t)$ and $\int_{\mathbb{R}} |f_n(t) - f_{\hat{\theta}}(t)|^p d\mu(t)$ can be used to test hypotheses of the forms $H_0 : f = f_0$, where f_0 is completely known, and $H_0 : f = f_{\theta}$, for a consistent estimator $\hat{\theta}$ of θ . When $p = 2$, Bickel and Rosenblatt (1973) show that whenever f satisfies certain smoothness conditions, one should reject H_0 for large values of the statistic $\int_{\mathbb{R}} |f_n(t) - f_{\hat{\theta}}(t)|^2 dt$, (based on the quantiles of the standard normal distribution), provided that $|\hat{\theta} - \theta_0| = o_p((nh_n)^{-1/2})$, where θ_0 is the true value of θ .

Unfortunately, as will be shown later in this article, simulation results confirm that the convergence of $I_n(p)$ (properly normalized) to the standard normal distribution can be rather slow. This fact was also noticed by Mojirsheibani (2007), who used Efron's original uniform bootstrap as an alternative approach to estimate the distribution of $I_n(p)$. In the next section we consider an unconditional weighted bootstrap approach to approximate the limiting distribution of $I_n(p)$. We also perform simulation studies to compare the performance of the proposed method, as compared to the classical CLT of Csörgő and Horváth (1988) and the original bootstrap of Efron (1979).

2 Main results

A number of authors have used the weighted bootstrap as a generalization of Efron (1979) original bootstrap in the literature. To the best of our knowledge, the first paper that introduced and used the concept of weighted bootstrap, with weights different from those of Efron in the cited paper, is Rubin (1981). His weighted bootstrap was applied under a Bayesian framework. Burke (2010) studies weighted empirical processes and their applications to change-point detection, whereas Burke (2000) uses Gaussian weights to form bootstrap confidence bands for a distribution function. Horváth et al. (2000) give the rate of the best Gaussian approximation for the weighted bootstrap empirical process and construct a sequence of Brownian bridges achieving this rate. A closely related result is given by Alvarez-Andrade and Bouzebda (2013). Mason and Newton (1992) give conditions under which the weighted bootstrapped mean is consistent. The monograph by Barbe and Bertail (1995) gives a general view and further results on the weighted bootstrap. Clearly Efron's original bootstrap itself is a weighted bootstrap algorithm, where the weights are multino-

mial random variables. One drawback of multinomial weights is that some observations may be sampled more than once while others are not sampled at all. Furthermore, in many applications, depending on the weights chosen, the weighted bootstrap has been shown to be computationally more efficient than Efron's algorithm; see, for example, (Burke, 2000; Hall and Mammen, 1994; Horváth et al., 2000).

In order to state our results, we first give a number of assumptions that will be used in the rest of this article. They are the same as those used in (Csörgő and Horváth, 1988).

Condition ($\mathcal{K}$) The kernel K satisfies

- (i) $\int K(t)dt = 1$, $\int tK(t)dt = 0$.
- (ii) There is a finite interval on which the kernel K is continuous and bounded, and vanishes outside of this interval.
- (iii) K is of bounded variation on the real line.

Csörgő and Horváth (1988) have also stated the additional condition that $\int K^2(t)dt < \infty$. This is condition C.8 in the cited paper; however, in view of (ii), this condition is unnecessary.

Conditions ($\mathcal{F}$)

- (i) $R(s) = \int f^{p/2}(t+s)d\mu(t)$ exists and is bounded in a neighborhood of zero.
- (ii) f is uniformly bounded, almost everywhere with respect to μ , and is monotone in a neighborhood of $t_F = \sup\{t : F(t) = 0\}$ and also in that of $t^F = \inf\{t : F(t) = 1\}$.
- (iii) $d\mu(t) = w(t)dt$ where the function $w(t) \geq 0$ is bounded and integrable over finite intervals, and $f^{p/2}(t)w(t)$ is uniformly continuous on the sets $\mathbb{R}_1^*, \dots, \mathbb{R}_k^*$, where the Lebesgue measure of $\mathbb{R} \setminus \bigcup_k \mathbb{R}_k^*$ is zero.
- (iv) The integral $\int [F(t+s)(1-F(t-s))]^{(0.5-\nu)p} w(t)dt$ exists and is bounded in a neighborhood of zero for some $0 < \nu \leq 0.5$.
- (v) For some $0 \leq \tau < \infty$, both $(|x|^\tau \vee 1)f''(x)$ and $(\log(|x|^\tau \vee e))^{1/2}|x|^\tau |f'(x)/f^{1/2}(x)|$ are a.e. uniformly bounded w.r.t. to μ and $\int (|x|^{-\tau} \wedge 1)^p w(t)dt < \infty$.

Using the notation of (Csörgő and Horváth, 1988), let

$$m = E|N|^p \left(\int K^2(t)dt \right)^{p/2} \int f^{(p+2)/2}(t)dt < \infty, \quad (1)$$

where $N = N(0, 1)$ stands for a standard normal random variable throughout this paper and

$$\sigma^2 = \sigma_1^2 \int f^{p+2}(t)dt \left(\int K^2(t)dt \right)^p, \quad (2)$$

where

$$\sigma_1^2 = \int \left(\left\{ (2\pi)^{-1} \int \int |xy|^p (1 - r^2(u))^{-1/2} \exp \left(-\frac{x^2 - 2r(u)xy + y^2}{2(1 - r^2(u))} \right) dx dy \right\} - (E|N|^p)^2 \right) du, \quad (3)$$

with

$$r(u) = \left(\int K^2(s) ds \right)^{-1} \int K(s) K(u + s) ds.$$

Let $0 < \nu \leq 1/2$ be as in condition $(\mathcal{F})(iv)$ and define $\kappa_\nu(n) = n^{-1/2} \log n$, if $\nu = 1/2$, and $\kappa_\nu(n) = n^{-\nu}$, if $0 < \nu < 1/2$. Then we have the following result ([Csörgő and Horváth, 1988](#)).

Theorem 1 *Suppose that conditions $(\mathcal{K})$ and $(\mathcal{F})$ hold with some $0 < \nu \leq 1/2$. If $nh_n^4 \rightarrow 0$ and $h_n^{-1} \kappa_\nu(n) \rightarrow 0$, as $n \rightarrow \infty$, then*

$$(h_n \sigma^2)^{-1/2} \{ (nh_n)^{p/2} I_n(p) - m \} \rightarrow^d N(0, 1), \quad 1 \leq p < \infty.$$

Next, to describe our proposed weighted bootstrap approach, let $\epsilon_1, \dots, \epsilon_n$ be i.i.d. random variables with mean $E(\epsilon_1)$ and variance 1, independent of $X_1, \dots, X_n$. The weighted bootstrap version of $f_n(x)$ is given by

$$f_{nn}^{(w)}(x) = (nh_n)^{-1} \sum_{i=1}^n W_{ni} K((x - X_i)/h_n), \quad \text{where } W_{ni} = 1 + \epsilon_i - \bar{\epsilon} \text{ and } \bar{\epsilon} = n^{-1} \sum_{i=1}^n \epsilon_i. \quad (4)$$

Similarly, define the weighted bootstrap statistic

$$I_{nn}^{(w)}(p) = \int |f_{nn}^{(w)}(x) - f_n(x)|^p f_n(x) dx, \quad 1 \leq p < \infty. \quad (5)$$

This is the weighted bootstrap version of $I_n(p) = \int_{\mathbb{R}} |f_n(t) - f(t)|^p f(t) dt$, i.e., the special important case where the function $w(t)$ in condition $(\mathcal{F})(iii)$ is equal to the unknown density function $f(t)$ itself. In passing we also note that if the terms W_{ni} in (4) are replaced by M_{ni} , $i = 1, \dots, n$, where $(M_{n1}, \dots, M_{nn})$ is a multinomial random vector on n categories (with probabilities $1/n$), then $f_{nn}^{(w)}(x)$ will coincide with Efron's original bootstrap estimate $f_{nn}(x) := (nh_n)^{-1} \sum_{i=1}^n K((x - X_i^*)/h_n)$, where $(X_1^*, \dots, X_n^*)$ is a random sample of size n drawn with replacement from $X_1, \dots, X_n$.

Next, to present our main results, put

$$\hat{m} = E|N|^p \left(\int K^2(t) dt \right)^{p/2} \int f_n^{(p+2)/2}(t) dt \quad (6)$$

and

$$\hat{\sigma}^2 = \sigma_1^2 \int f_n^{p+2}(t) dt \left(\int K^2(t) dt \right)^p, \quad (7)$$

where σ_1^2 is as in (3). In order to state our main results we also need to state the following conditions regarding the random variables (weights) $\epsilon_1, \dots, \epsilon_n$ used in (4).

Condition (M)

The random variables $\epsilon_1, \dots, \epsilon_n$ have a moment generating function in an open neighborhood of the origin.

Theorem 2 *Let $I_{nn}^{(w)}(p)$ be the weighted bootstrap L_p statistic in (5) and suppose that conditions (K) and (M) hold. Also, suppose that condition (F) holds when $w(t)$ is either 1 or $f(t)$, and that $f'(x)$ is uniformly bounded. If $(\log \log n)^{-1} nh_n^3 \rightarrow \infty$ and $nh_n^4 \rightarrow 0$, as $n \rightarrow \infty$, then*

$$(h_n \hat{\sigma}^2)^{-1/2} \{ (nh_n)^{p/2} I_{nn}^{(w)}(p) - \hat{m} \} \longrightarrow^d N(0, 1), \quad 1 \leq p < \infty.$$

Next, we consider the case where $w(t)$ is a known function of t . The weighted bootstrap counterpart of $I_{nn}^{(w)}(p)$ is then simply

$$J_{nn}^{(w)}(p) := \int |f_{nn}^{(w)}(x) - f_n(x)|^p w(x) dx, \quad 1 \leq p < \infty. \quad (8)$$

We then have the following result.

Theorem 3 *Let $J_{nn}^{(w)}(p)$ be the weighted bootstrap L_p statistic in (8) and suppose that conditions (K), (F), and (M) hold. If $(\log \log n)^{-1} nh_n^3 \rightarrow \infty$ and $nh_n^4 \rightarrow 0$, as $n \rightarrow \infty$, then*

$$(h_n \tilde{\sigma}^2)^{-1/2} \{ (nh_n)^{p/2} J_{nn}^{(w)}(p) - \tilde{m} \} \longrightarrow^d N(0, 1), \quad 1 \leq p < \infty,$$

where

$$\tilde{m} = E|N|^p \left(\int K^2(t) dt \right)^{p/2} \int f_n^{p/2}(t) w(t) dt$$

and

$$\tilde{\sigma}^2 = \left(\sigma_1^2 \int f_n^p(t) w^2(t) dt \right) \left(\int K^2(t) dt \right)^p,$$

with σ_1^2 as in (3).

3 Examples

In this section, we carry out some numerical studies to assess the performance of the following three methods: the CLT (Theorem 1), the proposed weighted bootstrap (Theorem 2), and Efron's original bootstrap. Our examples indicate that, in general, the weighted bootstrap can perform very well in terms of capturing the finite-sample distribution of $I_n(p)$. In the following examples, we consider random samples of sizes $n = 50$ and $n = 100$ drawn from the following mixture of normals:

$$f(x) = \frac{1}{3\sqrt{2\pi}}e^{-(x-2)^2/2} + \frac{2}{3\sqrt{2\pi}}e^{-x^2/2}.$$

For the kernel, we use the truncated Gaussian density

$$K(u) = \frac{I\{-5 < u < 5\}}{(\Phi(5) - \Phi(-5))\sqrt{2\pi}}e^{-u^2/2},$$

where Φ is the $N(0, 1)$ cumulative distribution function. For a sample of size n , we calculated the kernel density estimator of $f(x)$ and its two bootstrap versions (weighted and Efron's original bootstrap), as well as their corresponding L_2 -statistics. For the CLT, we calculated the kernel density estimate $f_n(x)$ over the grid of values of $h_n = 0.08, 0.1, 0.15, 0.25, 0.3, 0.35, 0.5, 0.7$, and 0.9 , as well as the corresponding statistic $I_n(2)$. For the weighted bootstrap, we generated a sample of weights $(\epsilon_1, \dots, \epsilon_n)$, where ϵ_i 's are i.i.d. Here, we considered two distributions for the weights: $N(0, 1)$ and $\text{Exp}(1)$. Then, we calculated the weighted bootstrap estimate $f_{nn}^{(w)}(x)$ for each value of h_n , as given by (4). This was repeated a total of $B = 1000$ times. The resulting weighted bootstrap statistics are $I_{nn,1}^{(w)}(2), \dots, I_{nn,B}^{(w)}(2)$, which were computed according to equation (5). Finally, for Efron's original bootstrap, we generated the bootstrap sample $(X_1^*, \dots, X_n^*)$, which was drawn with replacement from the original sample. The corresponding bootstrap kernel density estimator $f_{nn}(x) = (nh_n)^{-1} \sum_{i=1}^n K((x - X_i^*)/h_n)$ was then constructed for each value of h_n , which in turn produces the statistic $I_{nn}(2) = \int |f_{nn}(x) - f_n(x)|^2 f_n(x) dx$. Repeating this $B = 1000$ times will result in $I_{nn,1}(2), \dots, I_{nn,B}(2)$. Next, define

$$Z_n = \frac{h_n^{-1/2}[nh_n I_n(2) - \int K^2(t)dt \int f^2(t)dt]}{\sigma}, \quad (9)$$

$$Z_{nn,b} = \frac{h_n^{-1/2}[nh_n I_{nn,b}(2) - \int K^2(t)dt \int f_n^2(t)dt]}{\hat{\sigma}}, \quad b = 1, 2, \dots, B, \quad (10)$$

and

$$Z_{nn,b}^{(w)} = \frac{h_n^{-1/2}[nh_n I_{nn,b}^{(w)}(2) - \int K^2(t)dt \int f_n^2(t)dt]}{\hat{\sigma}}, \quad b = 1, 2, \dots, B, \quad (11)$$

where σ and $\hat{\sigma}$ are defined as in (2) and (7), respectively. Now observe that the asymptotic distributions of Z_n and $Z_{nn,b}^{(w)}$ are standard normal (by theorems 1 and 2). Furthermore, the asymptotic distribution of $Z_{nn,b}$ is also standard normal (Mojirsheibani, 2007). Therefore, if n is not “too small” then the random variables

$$W = \Phi(Z_n), \quad W_n = B^{-1} \sum_{b=1}^B I\{Z_{nn,b} \leq Z_n\}, \quad \text{and} \quad W_{nn} = B^{-1} \sum_{b=1}^B I\{Z_{nn,b}^{(w)} \leq Z_n\},$$

should have approximately a Unif(0,1) distribution. We carried out 400 such Monte Carlo runs which resulted in $W_1, \dots, W_{400}$, $W_{n,1}, \dots, W_{n,400}$, and $W_{nn,1}, \dots, W_{nn,400}$. Figure 1 presents the plots of the empirical distribution functions of W_i 's, $W_{n,i}$'s, and $W_{nn,i}$'s with sample size $n = 50$ for some values of h_n . The 45° line represents the true cdf of the Unif(0,1) distribution. Plots (d)-(f) show that the proposed weighted bootstrap, with N(0,1) weights, performs slightly better than Efron's original bootstrap (plots (g)-(i)) as well as the CLT approximation (plots (j)-(l)) in capturing the true distribution of $I_n(2)$. This is reflected by the fact that in plots (d)-(f), the empirical cdf of $W_{nn,i}$'s nearly coincide with the 45° line.

FIGURE 1 GOES HERE

Figure 2 gives the plots of the empirical distribution functions of W_i 's, $W_{n,i}$'s, and $W_{nn,i}$'s with the sample size $n = 100$. In this case, the weighted bootstrap with Exp(1) weights (plots (a)-(c)) clearly performs better than the other approximations.

FIGURE 2 GOES HERE

Next, we carried out formal tests of hypotheses of the distributions of the resulting 400 random variables. Two test statistics were employed: Kolmogorov-Smirnov, and Shapiro-Wilk tests. For each sample size n ($= 50$ or 100) and each choice of h_n ($= 0.08, 0.1, 0.15, 0.25, 0.3, 0.35, 0.5, 0.7$, and 0.9) the following 6 tests of the null hypothesis were carried out.

1. The Kolmogorov – Smirnov tests :	2. The Shapiro – Wilk tests (of normality) :
$H_0^{(1)} : W_1, \dots, W_{400}$ are iid Unif(0, 1)	$H_0^{(4)} : Z_{n,1}, \dots, Z_{n,400}$ are iid N(0, 1)
$H_0^{(2)} : W_{n,1}, \dots, W_{n,400}$ are iid Unif(0, 1)	$H_0^{(5)} : Z_{n,1}^B, \dots, Z_{n,400}^B$ are iid N(0, 1)
$H_0^{(3)} : W_{nn,1}, \dots, W_{nn,400}$ are iid Unif(0, 1).	$H_0^{(6)} : Z_{n,1}^w, \dots, Z_{n,400}^w$ are iid N(0, 1).

where $Z_{n,i}$'s are as before, $Z_n^B = \Phi^{-1}(W_n) = \Phi^{-1}[B^{-1} \sum_{b=1}^B I\{Z_{nn,b} \leq Z_n\}]$, and $Z_n^w = \Phi^{-1}(W_{nn}) = \Phi^{-1}[B^{-1} \sum_{b=1}^B I\{Z_{nn,b}^{(w)} \leq Z_n\}]$. For the CLT, the P -values corresponding to $H_0^{(1)}$ and $H_0^{(4)}$ were all less than 0.01 (and in most cases less than 10^{-5}), confirming the poor performance of the CLT of Theorem 1. On the other hand, for both Efron's and the proposed weighted bootstrap, all of the P -values corresponding to $H_0^{(2)}$, $H_0^{(3)}$, $H_0^{(5)}$, and $H_0^{(6)}$ were larger than 0.05, whenever $h_n < 0.5$. In particular, when $n = 50$, the P -values corresponding to $H_0^{(2)}$ were 0.3399, 0.1177, and 0.2809 for $h_n = 0.1, 0.15$, and 0.25 ,

respectively. The corresponding P -values for $H_0^{(3)}$, with $N(0,1)$ weights, were 0.7442, 0.2809, and 0.4067, respectively. Our results show that the test under $H_0^{(3)}$, with $N(0,1)$ weights, had larger P -values than the test under $H_0^{(2)}$ (Efron's bootstrap). A similar conclusion applies when $n = 100$: the test under $H_0^{(3)}$ with $\text{Exp}(1)$ weights had larger P -values than the test under $H_0^{(2)}$. In fact, the tests under $H_0^{(3)}$ and $H_0^{(6)}$ had larger P -values than tests under $H_0^{(2)}$ and $H_0^{(5)}$ for various choices of h_n .

4 Proofs

Proof of Theorem 2

To prove the theorem, we first state a number of technical lemmas. Without loss of generality, we will assume that all random variables and processes are defined on the same probability space (see, for example, section A.2 of Csörgő and Horváth (1993)). Define

$$F_n(x) = n^{-1} \sum_{i=1}^n \mathbf{I}\{X_i \leq x\}. \quad (12)$$

$$F_{nn}^{(w)}(x) = n^{-1} \sum_{i=1}^n (1 + \epsilon_i - \bar{\epsilon}) \mathbf{I}\{X_i \leq x\}. \quad (13)$$

$$\beta_n(x) = n^{1/2}(F_{nn}^{(w)}(x) - F_n(x)) = n^{-1/2} \sum_{i=1}^n (\epsilon_i - \bar{\epsilon}) \mathbf{I}\{X_i \leq x\}. \quad (14)$$

We first state a lemma which shows that the rate of convergence of the weighted bootstrap process $\beta_n(\cdot)$ to its limiting process, in the supnorm, is exactly the same as that of Komlós et al. (1975) approximation for the standard empirical process $\alpha_n(\cdot) = n^{1/2}(F_n(\cdot) - F(\cdot))$. More specifically, we have

Lemma 1 [Horváth et al. (2000; Theorem 1.3).]

Let $\beta_n(x)$ be as in (14). If the weights $\epsilon_1, \dots, \epsilon_n$ satisfy Condition $(\mathcal{M})$ then there is a sequence of Brownian bridges $\{B_n(t), 0 \leq t \leq 1\}$ such that

$$P \left\{ \sup_{-\infty < x < \infty} |\beta_n(x) - B_n(F(x))| > n^{-1/2}(c_1 \log n + y) \right\} \leq c_2 e^{-c_3 y},$$

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

In passing, we note that Lemma 1 in conjunction with the Borel-Cantelli lemma yields

$$\sup_{-\infty < y < \infty} |\beta_n(y) - B_n(F(y))| =^{a.s.} O(n^{-1/2} \log n). \quad (15)$$

To appreciate this, put $A_n(a) = \{(\log n)^{-1} \sqrt{n} \sup_{x \in \mathbb{R}} |\beta_n(x) - B_n(F(x))| > a\}$, for any a satisfying $a > c_1 + 1/c_3$. We need to show that $P\{\lim_{n \rightarrow \infty} A_n(a)\} = 0$. By the Borel-Cantelli lemma, it is sufficient to show that $\sum_{n=1}^{\infty} P\{A_n(a)\} < \infty$, for all $a > c_1 + 1/c_3$. Since $c_3(a - c_1) > 1$, it is straightforward to see that $\sum_{n=1}^{\infty} P\{A_n(a)\} \leq c_2 \sum_{n=1}^{\infty} n^{-c_3(a-c_1)} < \infty$.

Lemma 2 [Csörgő and Horváth (1988; Lemma 6).]

Let $\{W(t), 0 \leq t \leq 1\}$ be a Wiener process and put

$$\Gamma_n(t) = \int K\left(\frac{t-x}{h_n}\right) d(W(F(x)) - F(x)W(1)).$$

Also, suppose that conditions $(\mathcal{K})$ and $(\mathcal{F})(i), (ii), (iii),$ and (v) hold. If $h_n \rightarrow 0$, as $n \rightarrow \infty$, then

$$(h_n^{p+1} \sigma^2)^{-1/2} \left(\int |\Gamma_n(t)|^p w(t) dt - h_n^{p/2} m \right) \rightarrow^d N(0, 1), \quad 1 \leq p < \infty, \text{ as } n \rightarrow \infty,$$

where σ^2 and m are as in (2) and (1) respectively.

The proofs of the next three lemmas appear in the supplementary material file.

Lemma 3 Let $\{B_n(t), 0 \leq t \leq 1\}$, $n = 1, 2, \dots$ be the sequence of Brownian bridges of Lemma 1 and introduce

$$\Lambda_n(t) = \int K\left(\frac{t-x}{h_n}\right) dB_n(F(x)). \quad (16)$$

Then, under the conditions of Theorem 2 (excluding condition $(\mathcal{M})$), we have

$$Z_{n1} := (h_n^{p+1} \sigma^2)^{-1/2} \left(\int |\Lambda_n(t)|^p f_n(t) dt - h_n^{p/2} m \right) \rightarrow^d N(0, 1), \quad 1 \leq p < \infty, \text{ as } n \rightarrow \infty.$$

Lemma 4 Let $\Lambda_n(t)$ be as in (16), and define

$$\Delta_n = \int \left(\left| \int K\left(\frac{t-x}{h_n}\right) d\beta_n(x) \right|^p - |\Lambda_n(t)|^p \right) f_n(t) dt,$$

where $\beta_n(x)$ is the weighted bootstrap process in (14). Under the conditions of Theorem 2 (excluding condition $(\mathcal{M})$) we have

$$\Delta_n = O_p(n^{-1/2} h_n^{(p-1)/2} \log n), \quad \text{as } n \rightarrow \infty.$$

Lemma 5 Let the quantities m , σ^2 , $\hat{m}$, and $\hat{\sigma}^2$ be as in (1), (2), (6), and (7), respectively. Then, under the conditions of Theorem 2 (excluding condition $(\mathcal{M})$ and the boundedness of f'), $|\hat{m} - m|/\sqrt{h_n} \rightarrow^p 0$ and $\hat{\sigma}^2 - \sigma^2 \rightarrow^p 0$, as $n \rightarrow \infty$.

Proof of the Theorem

First note that

$$\begin{aligned}
I_{nn}^{(w)}(p) &= \int |f_{nn}^{(w)}(x) - f_n(x)|^p f_n(x) dx \\
&= \int h_n^{-p} \left| \int K\left(\frac{t-x}{h_n}\right) d(F_{nn}^{(w)}(x) - F_n(x)) \right|^p f_n(x) dx \\
&= n^{-p/2} h_n^{-p} \int \left| \int K\left(\frac{t-x}{h_n}\right) d\beta_n(x) \right|^p f_n(x) dx,
\end{aligned}$$

where $f_{nn}^{(w)}(x)$, $F_{nn}^{(w)}(x)$, and $\beta_n(x)$ are as in (4), (13), and (14), respectively. Therefore, we have

$$\begin{aligned}
&(h_n \sigma^2)^{-1/2} [(nh_n)^{p/2} I_{nn}^{(w)}(p) - \widehat{m}] \\
&= (h_n \sigma^2)^{-1/2} \left[h_n^{-p/2} \int |\Lambda_n(t)|^p f_n(t) dt - m \right] \\
&\quad + (h_n \sigma^2)^{-1/2} h_n^{-p/2} \int \left[\left| \int K\left(\frac{t-x}{h_n}\right) d\beta_n(x) \right|^p - |\Lambda_n(t)|^p \right] f_n(t) dt \\
&\quad + (h_n \sigma^2)^{-1/2} (m - \widehat{m}) \\
&:= Z_{n1} + Z_{n2} + Z_{n3}.
\end{aligned}$$

But, by Lemma 3, as $n \rightarrow \infty$

$$Z_{n1} = (h_n^{p+1} \sigma^2)^{-1/2} \left(\int |\Lambda_n(t)|^p f_n(t) dt - h_n^{p/2} m \right) \xrightarrow{d} N(0, 1), \quad 1 \leq p < \infty.$$

Furthermore, by Lemma 4

$$Z_{n2} = h_n^{-\frac{p+1}{2}} O_p \left(n^{-1/2} h_n^{\frac{p-1}{2}} \log n \right) = O_p \left(n^{-1/2} h_n^{-1} \log n \right) = o_p(1).$$

Similarly, by Lemma 5, as $n \rightarrow \infty$

$$Z_{n3} = o_p(1).$$

Since $\frac{\widehat{\sigma}^2}{\sigma^2} \xrightarrow{p} 1$ (Lemma 5), the theorem follows from an application of Slutsky's Theorem. $\square$

Proof of Theorem 3

The proof is similar to (and easier than) the proof of Theorem 2 and will not be given.

Discussion

In this article, we have proposed a weighted bootstrap method to approximate the distribution of L_p norms of kernel density estimators. We have also presented some numerical

evidence that suggests that the proposed method can perform well. As in most bootstrap procedures, the proposed method can be quite computer-intensive. Although our results in this paper have been obtained in the univariate framework, we believe that it is possible to extend them to general multivariate settings using the results of [Burke \(2010\)](#) on the generalization of Theorem 2.1 of [Horváth \(2000\)](#). In particular, our Lemma 1, which is an approximation result of Theorem 1.3 of [Horváth et al. \(2000\)](#), was itself proved using Theorem 2.1(i) of [Horváth \(2000\)](#), and this latter result is only suitable for the univariate setup. Theorem 2 of [Burke \(2010\)](#) makes it possible to establish a version of our Lemma 1 and, more importantly, a counterpart of our expression in (15), but with rates slower than $O(n^{-1/2} \log n)$. In fact, the rates will be of the form $O(n^{-\delta(d)} (\log n)^c)$, where d is the dimension of the multivariate vector $\mathbf{X}$. Here $0 < \delta(d) < 1/2$ and $c > 0$ are constants not depending on n . Obtaining the exact values of the constants $\delta(d)$ and c may be rather tedious.

One main drawback of any kernel type estimator is the choice of the bandwidth h_n . In the case of kernel density estimators, a popular choice is the one that minimizes the Asymptotic Mean Integrated Squared Error (AMISE) of the corresponding density estimator. Since AMISE depends on the unknown density functions, one can use the *plug-in* bandwidth selector of [Sheather and Jones \(1991\)](#), where $\hat{h}_n$ minimizes the plug-in estimate of AMISE. Extensions to multivariate density estimators have been developed by [Wand and Jones \(1994\)](#). Another method of finding a data-driven estimate of h_n is the *least-squares cross-validation* criterion of [Rudemo \(1982\)](#) and [Bowman \(1984\)](#). In our simulation studies, the main aim was simply to compare the 3 approximation methods, at different values of h_n , over 400 Monte Carlo runs. However, a small pilot study based on the plug-in estimate of the AMISE confirms that most of the selected values of $\hat{h}_n$ are in the range of 0.1 to 0.3, which is in line with the choices $h_n = 0.1, 0.15, 0.25$ used in figures 1 and 2. Therefore, if a choice such as $h_n = 0.15$ is good/poor for a particular data set (among 400 data sets), then it is good/poor for all three approximations because we are using the same kernel density estimate for each of the three approaches. Of course, in practice where there is just one data set (instead of 400), we must first find a good data-driven choice of h_n . As for the choice of the kernel K , condition (K) requires K to vanish outside of a compact set (for theoretical considerations). We believe that, from a numerical point of view, this condition does not play an important role, unless K has heavy tails.

Acknowledgements

This work is supported in part by the NSF Grant DMS-1407400 of Majid Mojirsheibani.

References

- Alvarez-Andrade, S. and Bouzebda, S. (2013). Strong approximations for weighted bootstrap of empirical and quantile processes with applications. *Stat. Methodol.*, 11:36–52.
- Barbe, P. and Bertail, P. (1995). *The weighted bootstrap, lecture notes in statistics*. Springer, Heidelberg.
- Beirlant, J. and Mason, D. (1995). On the asymptotic normality of l_p -norms of empirical functionals. *Math. Methods Statist.*, 4:1–19.
- Berlinet, A., Devroye, L., and Györfi, L. (1995). Asymptotic normality of l_1 -error in density estimation. *Statistics.*, 26:329–343.
- Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Stat.*, 1:1075–1095.
- Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71:353–360.
- Burke, M. (2000). Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. *Statist. Probab. Lett.*, 46:13–20.
- Burke, M. (2010). Approximations for a multivariate hybrid process with applications to change-point detection. *Math. Methods Statist.*, 19:121–135.
- Csörgő, M. and Horváth, L. (1988). Central limit theorems for l_p -norms of density estimators. *Probab. Theory Related Fields*, 80:269–291.
- Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley, Chichester.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7:1–26.
- Fryer, M. (1977). A review of some non-parametric methods of density estimation. *J. Inst. Math. Appl.*, 20:335–354.
- Hall, P. (1978). Limit theorems for stochastic measures of the accuracy of density estimators. *Stoch. Proc. Appl.*, 13:11–25.
- Hall, P. (1984). Central limit theorems for integrated square error of kernel density estimators. *J. Multivar. Anal.*, 14:1–16.
- Hall, P. and Mammen, E. (1994). On general resampling algorithms and their performance in distribution estimation. *Ann. Stat.*, 22:2011–2030.

- Horváth, L. (1991). On l_p -norms of multivariate density estimators. *Ann. Stat.*, 19:1933–1949.
- Horváth, L. (2000). Approximations for hybrids of empirical and partial sums processes. *J. Statist. Plann. Inference.*, 88:1–18.
- Horváth, L., Kokoszka, P., and Steinebach, J. (2000). Approximations for weighted bootstrap processes with an application. *Statist. Probab. Lett.*, 48:59–70.
- Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent random variables, and the sample distribution function. *Z. Wahrsch. Verw. Geb.*, 32:111–131.
- Mason, D. and Newton, M. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.*, 20:1611–1624.
- Mojirsheibani, M. (2007). Some approximations to l_p -statistics of kernel density estimators. *Statistics.*, 41:203–220.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.*, 33:1065–1076.
- Prakasa Rao, B. (1983). *Nonparametric functional estimation*. Academic Press, Orlando.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of density functions. *Ann. Math. Stat.*, 27:832–837.
- Rubin, D. (1981). The bayesian bootstrap. *Ann. Statist.*, 9:130–134.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Stat.*, 9:65–78.
- Sheather, S. and Jones, M. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. B*, 53:683–690.
- Steele, J. (1978). Invalidity of average squared error criterion in density estimation. *Canad. J. Statist.*, 6:193–200.
- Wand, M. and Jones, M. (1994). Multivariate plug-in bandwidth selection. *Comput. Stat.*, 9:97–116.
- Wegman, E. (1972). Nonparametric probability density estimation: A comparison of density estimation methods. *J. Statist. Comput. Simulation*, 1:225–245.

Figure1

(a) Weighted bootstrap with Exp(1) weights; ($h=0.1$)

(b) Weighted bootstrap with Exp(1) weights; ($h=0.15$)

(c) Weighted bootstrap with Exp(1) weights; ($h=0.25$)

(d) Weighted bootstrap with N(0,1) weights; ($h=0.1$)

(e) Weighted bootstrap with N(0,1) weights; ($h=0.15$)

(f) Weighted bootstrap with N(0,1) weights; ($h=0.25$)

(g) Efron's bootstrap ($h=0.1$)

(h) Efron's bootstrap ($h=0.15$)

(i) Efron's bootstrap ($h=0.25$)

(j) Large sample limit ($h=0.1$)

(k) Large sample limit ($h=0.15$)

(l) Large sample limit ($h=0.25$)

Figure2

(a) Weighted bootstrap with Exp(1) weights; (h=0.1)

(b) Weighted bootstrap with Exp(1) weights; (h=0.15)

(c) Weighted bootstrap with Exp(1) weights; (h=0.25)

(d) Weighted bootstrap with N(0,1) weights; (h=0.1)

(e) Weighted bootstrap with N(0,1) weights; (h=0.15)

(f) Weighted bootstrap with N(0,1) weights; (h=0.25)

(g) Efron's bootstrap (h=0.1)

(h) Efron's bootstrap (h=0.15)

(i) Efron's bootstrap (h=0.25)

(j) Large sample limit (h=0.1)

(k) Large sample limit (h=0.15)

(l) Large sample limit (h=0.25)