Weighted bootstrapped kernel density estimators in two-sample problems

Pouliot, William; Mojirsheibani, Majid

DOI:
10.1080/10485252.2016.1253842

Weighted bootstrapped kernel density estimators in two-sample problems

Majid Mojirsheibani[†1] **and William Pouliot**[††2]
Department of Mathematics, California State University, Northridge, CA 91330[†]
Department of Economics, University of Birmingham, Birmingham UK[††]

Forthcoming Journal of Non-Parametric Statistics

A weighted bootstrap method is proposed to approximate the distribution of the L_p $(1 \leq p < \infty)$ norms of two-sample statistics involving kernel density estimators. Using an approximation theorem of Horváth, Kokoszka and Steineback [(2000) 'Approximations for Weighted Bootstrap Processes with an Application', *Statistics and Probability Letters*, 48, 59-70], that allows one to replace the weighted bootstrap empirical process by a sequence of Gaussian processes, we establish an unconditional bootstrap central limit theorem for such statistics. The proposed method is quite straightforward to implement in practice. Furthermore, through some simulation studies, it will be shown that, depending on the weights chosen, the proposed weighted bootstrap approximation can sometimes outperform both the classical large-sample theory as well as Efron's [(1979) 'Bootstrap Methods: Another Look at the Jackknife', *Annals of Statistics*, 7, 1-26] original bootstrap algorithm.

Keywords: Kernel, density, weighted bootstrap, CLT, Brownian bridge.

Mathematics Subject Classification: 62G07; 62G08; 62G09

[1]Corresponding author. Email: majid.mojirsheibani@csun.edu
[2]Email: w.pouliot@bham.ac.uk

1 Introduction

Let $X_1, \cdots, X_n$ be independently and identically distributed (iid) random variables with probability density function (pdf) $f = F'$, where F is the cumulative distribution function (CDF) of X_1, and let $f_n(t) = (nh)^{-1} \sum_{i=1}^{n} K((t - X_i)/h)$ be the classical Parzen-Rosenblatt kernel estimators of f (Rosenblatt (1956) and Parzen (1962)). Here h is the *smoothing parameter* of the kernel K. A standard measure of the overall performance of f_n is given by the L_p-type statistic $I_n(p) = \int |f_n(t) - f(t)|^p \, d\mu(t), \ 1 \leq p < \infty$, where μ is a measure on the Borel sets of the real line.

Central limit theorems (CLT) for $I_n(p)$ have been available in the literature for more than four decades. These include the work of Bickel and Rosenblatt (1973) for the case of $p = 2$ and $d\mu(t) = a(t)dt$, where $a(t)$ is a bounded piece-wise smooth integrable function, and the work of Hall (1984) for the case of $p = 2$ and $\mu(t) = t$. CLT for $I_n(1)$ have been given by Berlinet, Devroye and Gyorfi (1995); also see Beirlant and Mason (1995). However, Csörgő and Horváth (1988), as well as Horváth (1991), established CLT for $I_n(p)$, for all values of $p \in [1, \infty)$ as well as more general measures μ. In fact, the important case of $\mu(t) = F(t)$ has also been of particular interest in the literature (Fryer (1977), Steele (1978), Hall (1982), Henze, Nikitin and Ebner (2009)). It is also well known that the convergence of $I_n(p)$ (properly scaled and normalized) to the standard normal distribution can be rather slow. See, for example, Mojirsheibani (2007) who suggests alternative methods to approximate the distribution of $I_n(p)$, based on Efron's original bootstrap.

The main goal of the present paper is to study and propose a weighted bootstrap approximation to distribution of the two-sample versions of $I_n(p)$. More specifically, we propose a weighted bootstrap approximation to the distribution of the statistic

$$I_{m,n}(p) \quad = \quad \int |f_{1,n}(t) - f_{2,m}(t)|^p \, d\mu(t), \qquad 1 \leq p < \infty \,, \tag{1}$$

where $f_{1,n}$ and $f_{2,m}$ are the kernel density estimators based on samples of sizes m and n, and are defined in (2), and where μ is a measure on the Borel sets of $\mathbb{R}$. Our numerical results show that the weighted bootstrap can often outperform Efron's original bootstrap. Furthermore, aside from its simplicity to use, in many applications the weighted bootstrap has been shown to be computationally more efficient than Efron's original algorithm (Burke (2000), Horváth et al. (2000) and Hall and Mammen (1994)). From a practical point of

view, the above statistics can be used to test the hypothesis $H_0 : f_1 = f_2$.

2 Main results

2.1 *Background*

Let $X_1, \cdots, X_n$ be iid random variables with the cdf F_1 and the pdf $f_1 = F_1'$. Also, let $Y_1, \cdots, Y_m$ be iid random variables with cdf F_2 and pdf $f_2 = F_2'$. Let $f_{1,n}$ and $f_{2,m}$ be the kernel estimators of f_1 and f_2, respectively, i.e.,

$$f_{1,n}(t) = (nh_1)^{-1} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_1}\right) \quad \text{and} \quad f_{2,m}(t) = (mh_2)^{-1} \sum_{i=1}^{m} K\left(\frac{t - Y_i}{h_2}\right). \tag{2}$$

Here the Kernel K is a real-valued function that satisfies certain regularity conditions (to be specified later) and $h_1 \equiv h_1(n)$ and $h_2 \equiv h_2(m)$ are smoothing parameters. Also let $I_{m,n}(p)$, $1 \leq p < \infty$, be the statistic introduced in (1). A central limit theorem due to Anderson, Hall and Titterington (1994) is available for the statistic in (1) for the case of $p = 2$ and $\mu(t) = t$. Closely relevant results are also given by Henze and Nikitin (2003). To address the general case where $p \in [1, \infty)$ and where $\mu(t)$ is not restricted to be $\mu(t) = t$, we first state a number of assumptions that will also be used later when we state our main results. These assumptions, which are the same as those used by Csörgő and Horváth (1988), are:

Condition $(\mathcal{K})$
The kernel K satisfies
(i) $\int K(t)dt = 1, \quad \int tK(t)dt = 0.$
(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.

Condition $(\mathcal{F})$
In what follows, f represents the pdf of X_1.
(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 bound, almost everywhere with respect to μ, and is monotone in a neighborhood of $t_{\mathrm{F}} = \sup\{t : \ F(t) = 0\}$ and also in that of $t^{\mathrm{F}} = \inf\{t : \ F(t) = 1\}$.
(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 $\mathbb{R}_1^*, \cdots, \mathbb{R}_k^*$, where the Lebesgue measure of $\mathbb{R} - \bigcup_k \mathbb{R}_k^*$ 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 with respect to μ and $\int (|x|^{-\tau} \wedge 1)^p w(x)dx < \infty$.

Also, define the quantities

$$\eta = E\,|N|^p \left(\int K^2(t)dt \right)^{p/2} \int f^{p/2}(t) d\mu(t) < \infty\,, \tag{3}$$

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

$$\sigma^2 = \sigma_1^2 \int f^p(t) \cdot w^2(t)dt \left(\int K^2(t)dt \right)^p\,, \tag{4}$$

where $w(t)$ is as in $\mathcal{F}(iii)$ and σ_1^2 is given by

$$\int \left(\left\{ (2\pi)^{-1} \int \int |xy|^p (1-r^2(u))^{-1/2} \exp\left(-\frac{x^2 - 2r(u)xy + y^2}{2(1-r^2(u))} \right) dxdy \right\} - (E|N|^p)^2 \right) du, \tag{5}$$

with

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

Now, let $\nu \in (0, 1/2]$ be as in $\mathcal{F}(iv)$ and define

$$\kappa_\nu(n) = \begin{cases} n^{-\nu} & \text{if} \quad 0 < \nu < 1/2 \\ n^{-\frac{1}{2}} \log n & \text{if} \quad \nu = 1/2 \end{cases} \tag{6}$$

Then the following result is the two-sample version of the classical CLT of Csörgő and Horváth (1988), provided that a common smoothing parameter is used for both $f_{1,n}$ and $f_{2,m}$, i.e., $h_1 = h_2 = h$.

Theorem 1 *Let $I_{m,n}(p)$, η, and σ^2 be as in (1), (3), and (4) respectively. Suppose that conditions $(\mathcal{K})$ and $(\mathcal{F})$ hold. If, as $m \wedge n \to \infty$,*

$$\frac{m}{n} \to c \in (0, \infty), \quad h \to 0, \quad \textit{and} \;\; h^{-1}\kappa_\nu(m \wedge n) \to 0\,,$$

where ν is as in (6), then

$$(h\sigma^2)^{-\frac{1}{2}} \left\{ \left(\frac{m+n}{mn} \right)^{-\frac{p}{2}} h^{\frac{p}{2}} \cdot I_{m,n}(p) - \eta \right\} \xrightarrow{d} N(0,1)\,, \quad 1 \leq p < \infty,$$

whenever $f_1 = f_2$.

Theorem 1 can be proved using weighted approximation techniques employed in Csörgő and Horváth (1988). In fact, for the sake of completeness, we have also provided a proof of Theorem 1 in the Appendix. However, our main aim is to propose a weighted bootstrap approximation of the distribution of $I_{m,n}(p)$. Although Theorem 1 holds for a large class of functions $w(t)$, where $w(t)$ is defined via $d\mu(t) = w(t)dt$ in condition $\mathcal{F}(iii)$, the most interesting choices are perhaps $w(t) = 1$ and $w(t) = f(t)$, where f is the common density under $H_0 : f_1 = f_2 \ (= f)$.

Our numerical studies in section 3 shows that in practice, and as in the one-sample problem, the convergence (to normality) in Theorem 1 can be rather poor. This fact remains true regardless of whether the parameters η and σ^2 are known (an unrealistic case) or they are estimated by some consistent estimators $\widehat{\eta}$ and $\widehat{\sigma}^2$. In the next section we propose a weighted bootstrap approximation to the distribution of $I_{m,n}(p)$, which works for all $p \in [1, \infty)$, is quite straightforward to implement, and has excellent finite sample performance as compared with both large-sample theory and Efron's (1979) original bootstrap algorithm.

2.2 *A weighted bootstrap approach*

In this section we consider a weighted bootstrap approximation of the distribution of $I_{m,n}(p)$. Several authors have used the weighted bootstrap as a generalization of Efron's (1979) original bootstrap in the literature. The first paper that used the concept of weighted bootstrap, with weights different from those of Efron in the cited paper, appears to be that of Rubin (1981). His weighted bootstrap was applied under a Bayesian framework. Burke (1998, 2000) uses a Gaussian weighted bootstrap to construct confidence bands for a cumulative distribution function. Burke (2010) studies weighted bootstrap empirical processes (in the framework of a hybrid empirical process) and discusses their applications to change-point detection. Horváth et al. (2000) study the rate of the best Gaussian approximation of the weighted bootstrap empirical process and construct a sequence of Brownian bridges achieving this rate. Alvarez-Andrade and Bouzebda (2013) consider strong approximations of weighted empirical and quantile processes and discuss the applications of their results to censored quantile processes. 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. We also note that Efron's original bootstrap is in fact a weighted bootstrap algorithm, where the weights are multi-

nomial random variables. However, depending on the weights used, in many applications the weighted bootstrap has been shown to be computationally more efficient than Efron's algorithm (Burke (2000), Horváth et al. (2000) and Hall and Mammen (1994))

The proposed weighted bootstrap approximation of $I_{m,n}(p)$ works as follows. Let $\delta_1, \cdots, \delta_n$ and $\delta'_1, \cdots, \delta'_m$ be iid random variables with mean $E(\delta_1)$ and variance 1, independent of the data $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$. The weighted bootstrap versions of the kernel density estimators $f_{1,n}$ and $f_{2,m}$ are, respectively,

$$\begin{aligned} f_{1,nn}(t) &= (nh)^{-1} \sum_{i=1}^{n} \left(1 + \delta_i - \bar{\delta}\right) K\left(\frac{t - X_i}{h}\right) \quad \text{and} \\ f_{2,mm}(t) &= (mh)^{-1} \sum_{i=1}^{m} \left(1 + \delta'_i - \bar{\delta}'\right) K\left(\frac{t - Y_i}{h}\right), \end{aligned}$$

where $\bar{\delta} = n^{-1} \sum_{i=1}^{n} \delta_i$ and $\bar{\delta}' = m^{-1} \sum_{i=1}^{m} \delta'_i$. To define the weighted bootstrap counterpart of (1), we first note that when $f_1 = f_2$, the statistic (1) can equivalently be written as

$$I_{m,n}(p) = \int |f_{1,n}(t) - f_{2,m}(t) - (f_1(t) - f_2(t))|^p \, w(t) dt \,, \tag{7}$$

where, in view of assumption $\mathcal{F}(iii)$, we have replaced $d\mu(t)$ by $w(t)dt$. Now (7) suggests considering the bootstrap statistic

$$\widehat{I}_{m,n}(p) = \int |f_{1,nn}(t) - f_{2,mm}(t) - (f_{1,n}(t) - f_{2,m}(t))|^p \, w(t) dt \,. \tag{8}$$

Clearly, the term $(f_{1,n}(t) - f_{2,m}(t))$ in (8), which is the kernel estimator of $f_1(t) - f_2(t) = 0$, is not necessarily zero and must be included. This is also in the spirit of the bootstrap hypothesis testing ideas discussed in Hall (1992; Sec. 3.12). In what follows, we will also pay particular attention to the special and important case where $w(t) = f(t)$. This choice of $w(t)$ has been of particular interest in the literature for general one- and two-sample problems; see, for example, Cao and van Keilegom (2006), Hall (1982), Steele (1978), Fryer (1977), Wegman (1972), and Rosenblatt (1952). In this case, the bootstrap counterpart of (7) is given by

$$\widetilde{I}_{m,n}(p) = \int |f_{1,nn}(t) - f_{2,mm}(t) - (f_{1,n}(t) - f_{2,m}(t))|^p \, f_{m+n}(t) dt \,, \tag{9}$$

where f_{m+n} is the "pooled" estimator of $f(t)$, i.e.,

$$\begin{aligned} f_{m+n}(t) &= \frac{1}{(n+m)h} \left[\sum_{i=1}^{n} K\left(\frac{t - X_i}{h}\right) + \sum_{i=1}^{m} K\left(\frac{t - Y_i}{h}\right) \right] \\ &= ((n+m)h)^{-1} \left[nh f_{1,n}(t) + mh f_{2,m}(t) \right] . \end{aligned}$$

To state our main results, we also need the following condition regarding the choice of the iid random variables $\delta_1, \cdots, \delta_n, \delta'_1, \cdots, \delta'_m$:

Condition $(\mathcal{M})$
The random variables $\delta_1, \cdots, \delta_n, \delta'_1, \cdots, \delta'_m$ are iid with finite mean $E(\delta_1)$ and variance 1, independent of $X_1, \cdots, X_n, Y_1, \cdots, Y_m$ and δ_1 has a moment generating function in an open neighborhood of the origin.

Now, let $\widetilde{\eta}$ and $\widetilde{\sigma}^2$ be the sample versions of η and σ^2 defined in (3) and (4), i.e.,

$$\widetilde{\eta} = E|N|^p \left(\int K^2(t)dt \right)^{p/2} \int f_{m+n}^{(p+2)/2}(t)dt \tag{10}$$

and

$$\widetilde{\sigma}^2 = \sigma_1^2 \int f_{m+n}^{p+2}(t)dt \left(\int K^2(t)dt \right)^p , \tag{11}$$

where σ_1^2 is as in (5).

Theorem 2 *Let $\widetilde{I}_{m,n}(p)$ be as in (9) and suppose that conditions $(\mathcal{K})$ and $(\mathcal{M})$ hold. Also suppose that condition $(\mathcal{F})$ holds when $w(t)$ is either 1 or $f(t)$. If, as $m \wedge n \to \infty$,*

$$m/n \to c \in (0, \infty), \quad h \to 0, \quad \text{and} \quad \frac{nh^3}{\log \log n} \to \infty ,$$

then

$$\left(h\widetilde{\sigma}^2\right)^{-\frac{1}{2}} \left\{ \left(\frac{m+n}{mn} \right)^{-\frac{p}{2}} h^{\frac{p}{2}} \cdot \widetilde{I}_{m,n}(p) - \widetilde{\eta} \right\} \xrightarrow{d} N(0,1) , \quad 1 \le p < \infty ,$$

whenever $f_1 = f_2$.

For the general case where $w(t)$ is a known function of t we have the following result.

Theorem 3 *Let $\widehat{I}_{m,n}(p)$ be as in (8) and suppose that conditions $(\mathcal{K})$, $(\mathcal{F})$, and $(\mathcal{M})$ hold. Also suppose that $\int d\mu(t) < \infty$. If, as $m \wedge n \to \infty$,*

$$m/n \to c \in (0, \infty), \ h \to 0, \ \text{and} \ nh^3 \to \infty ,$$

then

$$\left(h\widehat{\sigma}^2\right)^{-\frac{1}{2}} \left\{ \left(\frac{m+n}{mn} \right)^{-\frac{p}{2}} h^{\frac{p}{2}} \cdot \widehat{I}_{m,n}(p) - \widehat{\eta} \right\} \xrightarrow{d} N(0,1) , \quad 1 \le p < \infty ,$$

whenever $f_1 = f_2$, where the estimators $\widehat{\eta}$ and $\widehat{\sigma}^2$ are obtained from η and σ^2 upon replacing f by f_{m+n} in (3) and (4).

3 Numerical examples

In this section we perform some simulation studies to assess the finite-sample performance of the methods discussed this paper. These studies show that, in general, the proposed weighted bootstrap performs well in capturing the distribution of the statistics $I_{m,n}(p)$. Our numerical work involves samples of sizes $n = m = 40$, $n = m = 80$, and $n = 50$, $m = 15$ drawn from mixture of normals: $f(x) = \frac{1}{3}\phi(x-2)+\frac{2}{3}\phi(x)$ where $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ is the standard normal pdf. As for the kernel K, we considered a truncated Gaussian kernel given by $K(u) = c\phi(u)I\{-4 \le u \le 4\}$, where $c = (\Phi(4)-\Phi(-4))^{-1}$ is the norming constant, and Φ is the standard normal cdf. Next we computed the kernel density estimators $f_{1,n}$, $f_{2,m}$ as well as the corresponding two-sample statistic

$$Z_n := (h\sigma^2)^{-\frac{1}{2}}\left\{\left(\frac{m+n}{mn}\right)^{-\frac{p}{2}} h^{\frac{p}{2}} \cdot I_{m,n}(p) - \eta\right\}$$

for several values of the smoothing parameter h and $p = 2$; here we took $w(t) = f(t)$. Furthermore, for each pair of sample sizes (m, n) and each value of h, we computed the pooled estimator f_{m+n}, as well as 1000 copies of the weighted bootstrap density estimators $f_{1,nn}$ and $f_{2,mm}$ based on two different choices for the distribution of the weights $\delta_1, \cdots, \delta_n, \delta'_1, \cdots, \delta'_m$: N(0,1) and Exp(1). These 1000 values were then used to construct $B = 1000$ copies of the corresponding weighted bootstrap statistic (for each choice of the distribution of the weights)

$$Z_{nn} := \left(h\widetilde{\sigma}^2\right)^{-\frac{1}{2}}\left\{\left(\frac{m+n}{mn}\right)^{-\frac{p}{2}} h^{\frac{p}{2}} \cdot \widetilde{I}_{m,n}(p) - \widetilde{\eta}\right\},$$

where $\widetilde{I}_{m,n}(p)$ is as in (9). Furthermore, for each value of m, n, and h, we computed $B = 1000$ copies of

$$Z^*_n := \left(h\widetilde{\sigma}^2\right)^{-\frac{1}{2}}\left\{\left(\frac{m+n}{mn}\right)^{-\frac{p}{2}} h^{\frac{p}{2}} \cdot I^*_{m,n}(p) - \widetilde{\eta}\right\},$$

which are Efron's original bootstrap counterparts of Z_n, i.e.,

$$I^*_{m,n}(p) \;=\; \int \left|f^*_{1,n}(t) - f^*_{2,m}(t) - (f_{1,n}(t) - f_{2,m}(t))\right|^p f_{m+n}(t)dt\,.$$

Here $f^*_{1,n}(t) = (nh_1)^{-1}\sum_{i=1}^n K((t-X^*_i)/h)$ and $f^*_{2,m}(t) = (mh_2)^{-1}\sum_{i=1}^m K((t-Y^*_i)/h)$, where $X^*_1, \cdots, X^*_n$ is a sample of size n drawn with replacement from $X_1, \cdots, X_n$, and $Y^*_1, \cdots, Y^*_m$ is a sample of size m drawn with replacement from $Y_1, \cdots, Y_m$. Next observe that if m and n are not "too small" then by Theorem 1 the random variable

$$U := \Phi(Z_n)$$

should approximately have a Unif(0,1) distribution. Similarly, if we let $Z_{nn}(1), \cdots, Z_{nn}(1000)$, be the 1000 values of the weighted bootstrap statistic Z_{nn}, then by Theorem 2 the random variable

$$Y = B^{-1} \sum_{b=1}^{B} I\{Z_{nn}(b) \leq Z_n\}$$

is approximately a Unif(0,1) random variable. Likewise, if we let $Z_n^*(1), \cdots, Z_n^*(1000)$, represent the 1000 $(= B)$ values of Z_n^*, then

$$V = B^{-1} \sum_{b=1}^{B} I\{Z_n^*(b) \leq Z_n\}$$

is expected to be approximately a Unif(0,1) random variable. Repeating the entire above process a total of 400 times, we obtain $(U_1, \cdots, U_{400})$, $(Y_1, \cdots, Y_{400})$, and $(V_1, \cdots, V_{400})$.

[FIGURE 1 GOES HERE]

Figure 1 gives the plots of the empirical cdf of U_i's, Y_i's, and V_i's for various values of (m, n) and h=0.25 (c.f. Remark A for the choice of h). The 45° line represents the true cdf of the Unif(0,1) distribution. Plots (b), (c), and (d) show that when $n = m = 40$ the weighted bootstrap, as well as Efron's original bootstrap, perform substantially better than the large-sample theory approximation in plot (a). This is reflected by the fact that the empirical cdf of the V_i's in plot (b) and those of Y_i's in plots (c) and (d) nearly coincide with the 45° line. Furthermore, plot (d) shows that when Exp(1) weights are used, the weighted bootstrap is slightly superior to both N(0,1) weights and Efron's bootstrap. Similarly, when $n = m = 80$, both the weighted and Efron's bootstrap perform much better; see plots (e), (f),(g),(h). However, as plot (g) shows, in this case the weighted bootstrap with N(0,1) weights has a slight edge over Exp(1) weights. The case where $n = 50$ and $m = 15$ is given in plots (i), (j), (k), and (l). Once again the large-sample theory (plot (i)) gives poor results. In passing we also note that for this case the N(0,1) weights perform much better that the Exp(1) weights (compare plots (k) and (l)).

[FIGURE 2 GOES HERE]

Figure 2 gives the same plots for the case where $h = 0.4$ (c.f. Remark A). Once again plots (a)-(l) confirm that the weighted bootstrap can perform quite well.

Next, as a more formal approach, we have carried out tests of hypothesis for the distributions of the resulting random variables $(U_1, \cdots, U_{400})$, $(Y_1, \cdots, Y_{400})$, and $(V_1, \cdots, V_{400})$. We have employed two test statistics: Kolmogorov-Smirnov and Shapiro-Wilk tests. For each pair of sample sizes (m, n) and each of $h = 0.25$ and $h = .40$ the following tests of the null hypothesis were carried out.

1. The Kolmogorov-Smirnov tests:
$H_0^{(1)}$: $U_1, \cdots, U_{400}$ are iid Unif(0,1)
$H_0^{(2)}$: $Y_1, \cdots, Y_{400}$ are iid Unif(0,1) (using both N(0,1) and Exp(1) weights)
$H_0^{(3)}$: $V_1, \cdots, V_{400}$ are iid Unif(0,1)

2. The Shapiro-Wilk tests (of normality):
$H_0^{(4)}$: $\Phi^{-1}(U_1), \cdots, \Phi^{-1}(U_{400})$ are iid N(0,1)
$H_0^{(5)}$: $\Phi^{-1}(Y_1), \cdots, \Phi^{-1}(Y_{400})$ are iid N(0,1) (using both N(0,1) and Exp(1) weights)
$H_0^{(6)}$: $\Phi^{-1}(V_1), \cdots, \Phi^{-1}(V_{400})$ are iid N(0,1)

where Φ is the cdf of the standard normal distribution. A total of 24 tests were performed, corresponding to the 24 setups that gave rise to the 24 plots in Figures 1 and 2. The p-values corresponding to the hypotheses $H_0^{(k)}$, $k = 2, 3, 5, 6$ were all larger than 5% (and in fact, in most cases larger than 20%). This was true for all sample sizes and h values. On the other hand, all the other p-values (i.e., the p-values for $H_0^{(k)}$, $k = 1, 4$) were less than 10^{-10}, confirming the superior performance of the weighted bootstrap (as well as Efron's original bootstrap).

Remark A

For the choice of the bandwidth h in our simulation studies we have followed the approach used in Liu and Mojirsheibani (2015). More specifically, for kernel density estimators, a popular choice of h is the one that minimizes the Asymptotic Mean Integrated Squared Error (AMISE) of the corresponding density estimator. Since AMISE depends on the unknown density, one can use the plug-in bandwidth selector of Sheather and Jones (1991), where $\hat{h}$ minimizes the plug in estimate of AMISE. Extensions to multivariate density estimators have been developed by Wand and Jones (1994). Alternatively, one can find data-driven versions

of h by using the least-squares cross-validation criterion of Rudemo (1982) and Bowman (1984). For our simulation work, a preliminary small pilot study based on the plug-in estimate of the AMISE shows that the selected values of $\hat{h}$ are mainly in the range of 0.20 to 0.45, which justifies the choices $h = 0.25$ and 0.40 used in figures 1 and 2. We note that regardless of whether a choice, such as h = 0.25, is good/poor for a particular data set, it will be good/poor for each of the three approximations because we are using the same kernel density estimate for all three approaches

Remark B

In this paper we have established an unconditional bootstrap central limit theorem for the L_p norms of two-sample statistics involving kernel density estimators. In certain situations of practical interest both conditional and unconditional paradigms may be available; see, for example, Kosorok (2008, Sec. 10.1) where the weak convergence of weighted bootstrap empirical processes are addressed. In our setup we need more than just the weak convergence of such empirical processes; in fact, a key tool for establishing our main results is the strong approximation result of Horváth et al. (2000) which we have stated under Lemma 1. In the cited paper, these authors apply the weighted bootstrap method to problems in change-point analysis and establish the validity of their bootstrap approach in an unconditional sense; see Theorem 2.1 of Horváth et al. (2000). It appears that the strong approximation results for bootstrapped empirical processes established in Horváth et al. (2000), and substantially extended by Burke (2000), provide the right tools and the natural path for establishing unconditional bootstrap central limit theorems for many complicated statistics such as the ones we have addressed in this paper. Furthermore, although our unconditional bootstrap central limit theorems are stated in a univariate setting, we believe that with more efforts it is possible to extend them to a more general multivariate framework via Burke's (2010) generalization of Theorem 2.1 of Horváth (2000). More specifically, suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d $\mathbb{R}^d$-valued random vectors in an open neighborhood of the origin. If we put $\alpha_n(\mathbf{x}) = \sum_{i=1}^n \delta_i I\{\mathbf{X}_i \leq \mathbf{x}\}$ then by the results of Burke (2010: Theorem 2), there is a sequence of zero-mean Gaussian process, $W_m^F(\cdot)$, with covariance function $E(W^F(\mathbf{x})W^F(\mathbf{y})) = F(x_1 \wedge y_1, \ldots, x_d \wedge y_d)$, such that $P\{\sup_{\mathbf{x}\in\mathbb{R}^d} |n^{-1/2}\alpha_n(\mathbf{x}) - W_n^F(\mathbf{x})| \geq c_1 n^{-1/(2(2d-1))} \log n\} \leq c_2 n^{-2}$, where c_1 and c_2 are positive constants that depend on F and the distributuion of δ_1 only. Using this results in conjunction with Borel-Cantelli Lemma, Burke (2010) obtains the strong approximation

$\sup_{\mathbf{x}\in\mathbb{R}^d} |\alpha_n(\mathbf{x}) - W_n^F(\mathbf{x})| \stackrel{a.s.}{=} O(n^{-1/(2(2d-1))} \log n)$. These results are the multivariate versions of Theorem 2.1(i) of Horváth (2000, Theorem 1.3), which in turn is what we have stated under Lemma 1. Burke's (2010) results make it possible to retain a version of our Corollary 1 for the multivariate case, but with slower rates that depend on the dimension of d.

PROOF OF THEOREM 2

In what follows, we can and we will assume, without loss of generality, that all random variables and precesses are defined on the same probability space; for more information on this see section A.2 of Csörgő and Horváth (1993). To prove the theorem, we first state a KMT-type result of Horváth et al (2000) on the approximation of weighted bootstrap empirical processes. Let $\xi_1, \cdots, \xi_n$ be iid r.v.s with the cdf F. Let $F_n(t) = n^{-1} \sum_{i=1}^n I\{\xi_i \le t\}$, and $F_{nn}(t) = n^{-1} \sum_{i=1}^n (1 - \epsilon_i + \bar{\epsilon}) I\{\xi_i \le t\}$, where $\epsilon_1, \cdots, \epsilon_n$ are iid r.v.s with mean $E(\epsilon_1)$ and variance 1, and are independent of $\xi_1, \cdots, \xi_n$. Also, let $\beta_n(\cdot)$ be the weighted bootstrap empirical process, i.e., $\beta_n(t) = \sqrt{n}\,(F_{nn}(t) - F_n(t))\,,\ \ t \in \mathbb{R}\,.$

Lemma 1 *[Horváth et al (2000).] Let $\epsilon_1, \cdots, \epsilon_n$ be the iid r.v.s described above. Also, suppose that there is a $t_0 > 0$ such that $E(e^{t\epsilon_1}) < \infty$ for all $t \in (-t_0, t_0)$. Then there exists a sequence of Brownian bridges $\{B_n(t)\,,\ 0 \le t \le 1\}_{n\ge1}$ such that*

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

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

An immediate consequence of the above result is the following corollary:

Corollary 1 *Let $\epsilon_1, \cdots, \epsilon_n, \beta_n(\cdot)$, and $B_n(F(\cdot))$ be as in Lemma 1. Then*

$$\sup_{-\infty<t<\infty} |\beta_n(t) - B_n(F(t))| \ \stackrel{a.s.}{=}\ \mathcal{O}\left(n^{-1/2} \log n\right)\,.$$

We also make use of the following technical result.

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

$$\Gamma_n(t) = \int K\left(\frac{t-x}{h}\right) d(W(F(x)) - F(x)W(1)), \tag{12}$$

where $W(\cdot)$ is a standard Wiener process. If $h \to 0$ (as $n \to \infty$) then under conditions $(\mathcal{K})$

and $(\mathcal{F})(i),(ii),(iii),(v)$,

$$(h^{p+1}\sigma^2)^{-\frac{1}{2}}\left\{\int|\Gamma_n(t)|^p\,w(t)dt-h^{\frac{p}{2}}\eta\right\}\xrightarrow{d}N(0,1),$$

as $n\to\infty$, *where* η *are* σ^2 *are as in* (3) *and* (4), *respectively.*

Define

$$\begin{aligned}
F_{1,nn}(t) &= n^{-1}\sum_{i=1}^{n}\left(1-\delta_i+\bar{\delta}\right)I\left\{X_i\le t\right\}\\
F_{2,mm}(t) &= m^{-1}\sum_{i=1}^{m}\left(1-\delta_i'+\bar{\delta}'\right)I\left\{Y_i\le t\right\}\\
\beta_n^{(1)}(t) &= \sqrt{n}\left(F_{1,nn}(t)-F_{1,n}(t)\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\delta_i-\bar{\delta}\right)I\left\{X_i\le t\right\}\\
\beta_m^{(2)}(t) &= \sqrt{m}\left(F_{2,mm}(t)-F_{2,m}(t)\right)=\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\left(\delta_i'-\bar{\delta}'\right)I\left\{Y_i\le t\right\}\ .
\end{aligned}$$

Now, using the fact that

$$\begin{aligned}
f_{1,nn}(t)-f_{1,n}(t) &= h^{-1}\int K\left(\frac{t-x}{h}\right)d\left(F_{1,nn}(x)-F_{1,n}(x)\right)\\
f_{2,mm}(t)-f_{2,m}(t) &= h^{-1}\int K\left(\frac{t-x}{h}\right)d\left(F_{2,mm}(x)-F_{2,m}(x)\right)\ ,
\end{aligned}$$

we can write

$$\begin{aligned}
\widetilde{I}_{m,n}(p) &= \int\left|f_{1,nn}(t)-f_{2,mm}(t)-(f_{1,n}(t)-f_{2,m}(t))\right|^p f_{m+n}(t)dt\\
&= \left(\frac{m+n}{mn}\right)^{\frac{p}{2}}h^{-p}\int\left|\int K\left(\frac{t-x}{h}\right)\sqrt{\frac{mn}{m+n}}d\left\{(F_{1,nn}(x)-F_{1,n}(x))\right.\right.\\
&\quad\left.\left.+\,(F_{2,mm}(x)-F_{2,m}(x))\right\}\right|^p f_{m+n}(t)dt\\
&= \left(\frac{m+n}{mn}\right)^{\frac{p}{2}}h^{-p}\int\left|\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}\beta_n^{(1)}(x)-\sqrt{\frac{n}{m+n}}\beta_m^{(2)}(x)\right\}\right|^p f_{m+n}(t)dt\,.
\end{aligned}$$

Furthermore, by Lemma 1 and Corollary 1, there are two independent sequences of Brownian bridges $\{B_n^{(1)}(t),0\le t\le 1\}_{n\ge1}$ and $\{B_m^{(2)}(t),0\le t\le 1\}_{m\ge1}$ such that

$$\begin{aligned}
&\sup_{-\infty<t<\infty}\left|\beta_n^{(1)}(t)-B_n^{(1)}(F_1(t))\right| \overset{\text{a.s.}}{=} \mathcal{O}\left(n^{-1/2}\log n\right)\\
\text{and}\quad &\sup_{-\infty<t<\infty}\left|\beta_m^{(2)}(t)-B_m^{(2)}(F_2(t))\right| \overset{\text{a.s.}}{=} \mathcal{O}\left(m^{-1/2}\log m\right)\ .
\end{aligned}$$

Therefore, with $B_n^{(1)}(\cdot)$ and $B_m^{(2)}(\cdot)$ as above, one finds

$$
\begin{aligned}
&\left(\frac{m+n}{mn}\right)^{-\frac{p}{2}} h^p \cdot \widetilde{I}_{m,n}(p) \\
&= \int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x))\right\}\right|^p f_{m+n}(t)dt \\
&\quad + \left[\int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} \beta_n^{(1)}(x) - \sqrt{\frac{n}{m+n}} \beta_m^{(2)}(x)\right\}\right|^p f_{m+n}(t)dt \right. \\
&\qquad \left. - \int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x))\right\}\right|^p f_{m+n}(t)dt \right] \\
&:= S_{m,n} + T_{m,n}\,. \qquad (13)
\end{aligned}
$$

Using the inequality $||a|^p - |b|^p| \le p2^{p-1}|a-b|^p + p2^{p-1}|b|^{p-1}|a-b|$, where $p \ge 1$, one finds

$$
\begin{aligned}
&\frac{1}{p2^{p-1}} |T_{m,n}| \\
&\le \left[\int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} \left(\beta_n^{(1)}(x) - B_n^{(1)}(F(x))\right)\right.\right. \\
&\qquad\qquad \left.\left. - \sqrt{\frac{n}{m+n}} \left(\beta_m^{(2)}(x) - B_m^{(2)}(F(x))\right)\right\}\right|^p f_{m+n}(t)dt \right] \\
&\quad + \left[\int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x))\right\}\right|^{p-1} \right. \\
&\qquad\qquad \times \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} \left(\beta_n^{(1)}(x) - B_n^{(1)}(F(x))\right)\right.\right. \\
&\qquad\qquad\qquad \left.\left. \left. - \sqrt{\frac{n}{m+n}} \left(\beta_m^{(2)}(x) - B_m^{(2)}(F(x))\right)\right\}\right| f_{m+n}(t)dt \right] \\
&:= [T_{m,n}(i)] + [T_{m,n}(ii)]\,. \qquad (14)
\end{aligned}
$$

Since $|x|^p$ is a convex function of x for $p \ge 1$, we have the elementary inequality (also known as Loeve's inequality) that $|(x+y)/2|^p \le (|x|^p + |y|^p)/2\,, \quad p \ge 1$. Using this, we find

$$
\begin{aligned}
T_{m,n}(i) &\le 2^{p-1} \int \left| \int K\left(\frac{t-x}{h}\right) \sqrt{\frac{m}{m+n}} d\left\{\beta_n^{(1)}(x) - B_n^{(1)}(F(x))\right\}\right|^p f_{m+n}(t)dt \\
&\quad + 2^{p-1} \int \left| \int K\left(\frac{t-x}{h}\right) \sqrt{\frac{n}{m+n}} d\left\{\beta_m^{(2)}(x) - B_m^{(2)}(F(x))\right\}\right|^p f_{m+n}(t)dt \\
&:= 2^{p-1} T_{m,n,1}(i) + 2^{p-1} T_{m,n,2}(i)\,. \qquad (15)
\end{aligned}
$$

Now,

$$\begin{aligned}
&\left|\int K\left(\frac{t-x}{h}\right)\sqrt{\frac{m}{m+n}}d\left\{\beta_n^{(1)}(x)-B_n^{(1)}(F(x))\right\}\right|^p \\
&\quad=\left(\frac{m}{m+n}\right)^{\frac{p}{2}}\left|\int\left[\beta_n^{(1)}(t-hy)-B_n^{(1)}(F(t-hy))\right]dK(y)\right|^p \\
&\quad\le\left(\frac{m}{m+n}\right)^{\frac{p}{2}}\left(\sup_{u\in\mathbb{R}}\left|\beta_n^{(1)}(u)-B_n^{(1)}(F(u))\right|\cdot\int dK(y)\right)^p \\
&\quad\overset{\text{a.s.}}{=}\mathcal{O}(1)\cdot\mathcal{O}\left(\left(n^{-1/2}\log n\right)^p\right),\quad\text{(by Corollary 1)},
\end{aligned}\tag{16}$$

where the $\mathcal{O}(1)$ term follows from the fact that $\frac{m}{m+n}=\mathcal{O}(1)$, (because $\frac{m}{n}\to c\in(0,\infty)$). Since the $\mathcal{O}((n^{-1/2}\log n)^p)$ term in (16) does not depend on t, we find

$$T_{m,n,1}(i)\ \overset{\text{a.s.}}{\le}\ \mathcal{O}\left(\left(n^{-1/2}\log n\right)^p\right)\cdot\int f_{m+n}(t)dt\ =\ \mathcal{O}\left(\left(n^{-1/2}\log n\right)^p\right).$$

Similarly, $T_{m,n,2}(i)\overset{\text{a.s.}}{=}\mathcal{O}((m^{-1/2}\log m)^p)$. Therefore,

$$T_{m,n}(i)\ \overset{\text{a.s.}}{=}\ \mathcal{O}\left((n^{-\frac{1}{2}}\log n)^p\vee(m^{-\frac{1}{2}}\log m)^p\right).\tag{17}$$

Next, to deal with the term $T_{m,n}(ii)$ in (14), we use Hölder's inequality to find

$$\begin{aligned}
T_{m,n}(ii)\ \le\ &\left(\int\left|\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))-\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\}\right|^p f_{m+n}(t)dt\right)^{\frac{p-1}{p}} \\
&\times\left(\int\left|\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}\left(\beta_n^{(1)}(x)-B_n^{(1)}(F(x))\right)\right.\right.\right. \\
&\qquad\left.\left.\left.-\sqrt{\frac{n}{m+n}}\left(\beta_m^{(2)}(x)-B_m^{(2)}(F(x))\right)\right\}\right|^p f_{m+n}(t)dt\right)^{\frac{1}{p}} \\
=\ &\left(\int\left|\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))-\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\}\right|^p f_{m+n}(t)dt\right)^{\frac{p-1}{p}} \\
&\times\left(T_{m,n}(i)\right)^{\frac{1}{p}}
\end{aligned}\tag{18}$$

Now, let $\{B^{(1)}(t),0\le t\le 1\}$ and $\{B^{(2)}(t),0\le t\le 1\}$ be independent Brownian bridges and note that for each $m=1,2,\cdots$, and $n=1,2,\cdots$,

$$\begin{aligned}
&\left\{\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))-\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\},\quad t\in\mathbb{R}\right\} \\
&\quad\overset{\text{d}}{=}\left\{\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}B^{(1)}(F(x))-\sqrt{\frac{n}{m+n}}B^{(2)}(F(x))\right\},\quad t\in\mathbb{R}\right\}.
\end{aligned}\tag{19}$$

But, by the standard properties of independent Brownian bridges (see, for example, Shorack and Wellner (1986; page 32)), the process

$$\left\{ \sqrt{\frac{m}{m+n}} B^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B^{(2)}(F(x)) \,, \quad x \in \mathbb{R} \right\} \tag{20}$$

is also a Brownian bridge on $[0,1]$. Now, let

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

where $W(\cdot)$ is a standard Wiener process, and note that by Lemma 2 (with $w(t) = f(t)$)

$$\int |\Gamma_n(t)|^p \, f(t)dt \;=\; \mathcal{O}_p(h^{\frac{p}{2}}) \tag{21}$$

Furthermore, $\{W(s) - sW(1)\,, \;\; 0 \le s \le 1\} \stackrel{\mathrm{d}}{=} \{B(s)\,, \;\; 0 \le s \le 1\}$, where $B(\cdot)$ is a Brownian bridge. Thus, in view of (19) and (21), and the fact that (20) is a Brownian bridge on $[0\,,1]$, we obtain

$$\begin{aligned} &\int \left| \int K\left(\frac{t-x}{h}\right) d\left\{ \sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x)) \right\} \right|^p f(t)dt \\ &= \mathcal{O}_p(h^{\frac{p}{2}}) \,. \end{aligned} \tag{22}$$

Similarly, taking $w(t) = 1$ in Lemma 2, we have $\int |\Gamma_n(t)|^p dt = \mathcal{O}_p(h^{p/2})$, and therefore in view of (19)

$$\int \left| \int K\left(\frac{t-x}{h}\right) d\left\{ \sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x)) \right\} \right|^p dt \;=\; \mathcal{O}_p(h^{\frac{p}{2}}) \,. \tag{23}$$

Consequently,

$$\begin{aligned} &\int \left| \int K\left(\frac{t-x}{h}\right) d\left\{ \sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x)) \right\} \right|^p f_{m+n}(t)dt \\ &\le \int \left| \int K\left(\frac{t-x}{h}\right) d\left\{ \sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x)) \right\} \right|^p f(t)dt \\ &\quad + \sup_t \Big| f_{m+n}(t) - f(t) \Big| \\ &\qquad \times \int \left| \int K\left(\frac{t-x}{h}\right) d\left\{ \sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x)) \right\} \right|^p dt \\ &= \mathcal{O}_p\left(h^{\frac{p}{2}}\right) + \left\{ \mathcal{O}_p\big((m+n)^{-\frac{1}{2}} h^{-1} (\log\log(m+n))^{\frac{1}{2}}\big) + \mathcal{O}(h) \right\} \times \mathcal{O}_p\left(h^{\frac{p}{2}}\right) \\ &= \mathcal{O}_p\left(h^{\frac{p}{2}}\right) \,, \end{aligned} \tag{24}$$

where we have used the following classical result on the uniform performance of kernel density estimators (see, for example, Theorem 3.1.12 of Prakasa Rao (1983))

$$\begin{aligned}\sup_t |f_{m+n}(t) - f(t)| &\leq \sup_t |f_{m+n}(t) - E(f_{m+n}(t))| + \mathcal{O}(h) \\ &\overset{\text{a.s.}}{=} \mathcal{O}\left((m+n)^{-\frac{1}{2}} h^{-1} (\log\log(m+n))^{1/2}\right) + \mathcal{O}(h)\,. \end{aligned} \tag{25}$$

Now, by (14), (17), (18), and (24), we have

$$T_{m,n}(ii) = \mathcal{O}_p\left(h^{\frac{p-1}{2}}\left((n^{-\frac{1}{2}}\log n) \vee (m^{-\frac{1}{2}}\log m)\right)\right)\,. \tag{26}$$

Finally, putting together (26), (17), (14) we find

$$|T_{m,n}| = \mathcal{O}_p\left((n^{-\frac{1}{2}}\log n)^p \vee (m^{-\frac{1}{2}}\log m)^p\right) + \mathcal{O}_p\left(h^{\frac{p-1}{2}}\left((n^{-\frac{1}{2}}\log n) \vee (m^{-\frac{1}{2}}\log m)\right)\right)\,. \tag{27}$$

Now, observe that by (13)

$$\begin{aligned}
&\left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}\left\{\left(\frac{m+n}{mn}\right)^{-\frac{p}{2}} h^p \cdot \widetilde{I}_{m,n}(p) - h^{\frac{p}{2}}\widetilde{\eta}\right\} \\
&\overset{\text{by (13)}}{=} \left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}\left\{S_{m,n} + T_{m,n} - h^{\frac{p}{2}}\widetilde{\eta}\right\} \\
&= \left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}\left\{S_{m,n} - h^{\frac{p}{2}}\eta\right\} + \left(h\sigma^2\right)^{-\frac{1}{2}} \cdot (\widetilde{\eta} - \eta) + \left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}} \cdot T_{m,n} \\
&= \left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}\left[\int\left|\int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\}\right|^p \right. \\
&\qquad\qquad \left. \times f(t)dt - h^{\frac{p}{2}}\eta\right] \\
&\quad + \left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}\left[\int\left|\int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\}\right|^p \right. \\
&\qquad\qquad \left. \times \left(f_{m+n}(t) - f(t)\right) dt\right] \\
&\quad + \left(h\sigma^2\right)^{-\frac{1}{2}} \cdot (\widetilde{\eta} - \eta) \\
&\quad + \left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}} \cdot T_{m,n} \\
&:= \Delta_{m,n}(1) + \Delta_{m,n}(2) + \Delta_{m,n}(3) + \Delta_{m,n}(4)\,.
\end{aligned} \tag{28}$$

But by (27)

$$|\Delta_{m,n}(4)| = h^{-\frac{p+1}{2}} h^{\frac{p-1}{2}}\left((n^{-\frac{1}{2}}\log n) \vee (m^{-\frac{1}{2}}\log m)\right) \cdot \mathcal{O}_p(1) = o_p(1)\,. \tag{29}$$

It will also be shown in the Appendix that

$$h^{-\frac{1}{2}}\,|\widetilde{\eta}-\eta| \xrightarrow{\text{p}} 0 \;\text{ and }\; \widetilde{\sigma}^2 \xrightarrow{\text{p}} \sigma^2\,. \tag{30}$$

Thus, by (30),

$$|\Delta_{m,n}(3)| = o_p(1)\,. \tag{31}$$

Next, the independence of the sequences of Brownian bridges $B_n^{(1)}(\cdot),\ n\geq 1$ and $B_m^{(2)}(\cdot),\ m\geq 1$ in conjunction with (19) and Lemma 2 (with $w(t)=f(t)$) imply

$$\Delta_{m,n}(1) \xrightarrow{\text{d}} N(0,1)\,. \tag{32}$$

As for the term $\Delta_{m,n}(2)$, (23) and the arguments that lead to (24) yields

$$\begin{aligned}
\big|\Delta_{m,n}(2)\big| &\leq \big(h^{p+1}\sigma^2\big)^{-\frac{1}{2}}\cdot\sup_t\big|f_{m+n}(t)-f(t)\big| \\
&\quad\times\int\left|\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))-\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\}\right|^p dt \\
&= \big(h^{p+1}\sigma^2\big)^{-\frac{1}{2}}\cdot\Big\{\mathcal{O}_p\big(h^{-1}(m+n)^{-\frac{1}{2}}\,(\log\log(m+n))^{\frac{1}{2}}\,\big)+\mathcal{O}(h)\Big\}\cdot\mathcal{O}_p\left(h^{\frac{p}{2}}\right) \\
&= \mathcal{O}_p\big(h^{-\frac{3}{2}}(m+n)^{-\frac{1}{2}}\,(\log\log(m+n))^{\frac{1}{2}}\,\big) \\
&= o_p(1)\,, \qquad (\text{since } (\log\log n)^{-1}(n+m)h^3\to\infty\,,\ \text{as } m,n\to\infty)\,.
\end{aligned}$$

Thus,

$$\big(h^{p+1}\sigma^2\big)^{-\frac{1}{2}}\left\{\left(\frac{m+n}{mn}\right)^{-\frac{p}{2}}h^p\cdot\widetilde{I}_{m,n}(p)-h^{\frac{p}{2}}\widetilde{\eta}\right\}\xrightarrow{\text{d}} N(0,1)\,.$$

The proof of the theorem now follows from an application of Slutsky's theorem (since $\frac{\widetilde{\sigma}^2}{\sigma^2}\xrightarrow{\text{p}}1$).

□

PROOF OF THEOREM 3

The proof of this theorem is the same as (and in fact easier than) the proof of Theorem 2 and will not be given.

Appendix

PROOF OF (30)

We will actually prove the stronger result that $h^{-1/2}\,|\widetilde{\eta}-\eta| \stackrel{\text{a.s.}}{\to} 0$ and $\widetilde{\sigma}^2 \stackrel{\text{a.s.}}{\to} \sigma^2$. First note that

$$\begin{aligned}
|\hat{\eta}-\eta| &\leq c_1 \int \left| f_{m+n}^{(p+2)/2}(t) - f^{(p+2)/2}(t) \right| dt \\
&\qquad \text{(where the constant } c_1 \text{ does not depend on } f \text{ or } n\text{)} \\
&\leq c_2 \int |f_{m+n}(t) - f(t)|^{(p+2)/2}\, dt \\
&\qquad + \left(\int f^{(p+2)/2}(t)\, dt \right)^{1-2/(p+2)} \left(\int |f_{m+n}(t) - f(t)|^{(p+2)/2} dt \right)^{2/(p+2)} ,
\end{aligned}$$

where $c_2 = ((p+2)/2)2^{(p+2)/2-1} \cdot c_1$. Therefore, to show $h^{-1/2}|\widetilde{\eta}-\eta| \stackrel{\text{a.s.}}{\to} 0$, it is sufficient to show that $h^{-1/2}\big(\int |f_{m+n}(t) - f(t)|^{(p+2)/2} dt\big)^{2/(p+2)} \stackrel{\text{a.s.}}{\to} 0$. Now observe that

$$\begin{aligned}
\int |f_{m+n}(t) - f(t)|^{(p+2)/2} dt &\leq \left[\sup_t |f_{m+n}(t) - f(t)| \right]^{p/2} \times \int |f_{m+n}(t) - f(t)| dt \\
&:= \xi_{m+n}^{(1)} \times \xi_{m+n}^{(2)} .
\end{aligned}$$

But, by (25), $\xi_{m+n}^{(1)} \stackrel{\text{a.s.}}{=} \mathcal{O}(((m+n)^{-\frac{1}{2}} h^{-1} (\log\log(m+n))^{1/2} \vee h)^{p/2})$. Furthermore, by the well-known results on the almost-sure behavior of L_1-norms of kernel density estimators (see Eggermont and Lariccia (2001; page 149)) we have

$$\xi_{m+n}^{(2)} \stackrel{\text{a.s.}}{=} \mathcal{O}\big(h^2 + ((m+n)h)^{-1/2} + ((m+n)^{-1} \log(m+n))^{1/2}\big) .$$

It is now quite straightforward to verify that $h^{-1/2}\big(\xi_{m+n}^{(1)} \cdot \xi_{m+n}^{(2)}\big)^{2/(p+2)} \stackrel{\text{a.s.}}{\to} 0$. The proof of $\widetilde{\sigma}^2 \stackrel{\text{a.s.}}{\to} \sigma^2$ is virtually the same (and in fact easier) and will not be given.

PROOF OF THEOREM 1

Let $X_1, \cdots, X_n \stackrel{\text{iid}}{\sim} f_1$ and $Y_1, \cdots, Y_m \stackrel{\text{iid}}{\sim} f_2$. Define the empirical distribution functions

$$F_{1,n}(t) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq t\} \quad \text{and} \quad F_{2,m}(t) = \frac{1}{m} \sum_{i=1}^{m} I\{Y_i \leq t\} ,$$

and let $\alpha_n^{(1)}(\cdot)$ and $\alpha_m^{(2)}(\cdot)$ be the corresponding empirical processes, i.e.,

$$\alpha_n^{(1)}(\cdot) = \sqrt{n}\,(F_{1,n}(t) - F_1(t)) \quad \text{and} \quad \alpha_m^{(2)}(\cdot) = \sqrt{m}\,(F_{2,m}(t) - F_2(t)) .$$

Now, observe that

$$\begin{aligned} I_{m,n}(p) &= \int h^{-p} \left| \int K\left(\frac{t-x}{h}\right) d\left(F_{1,n}(x) - F_{2,m}(x)\right) \right|^p w(t)dt \\ &= \left(\frac{m+n}{mn}\right)^{\frac{p}{2}} h^{-p} \int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{mn}{m+n}}\left(F_{1,n}(x) - F_{2,m}(x)\right)\right\} \right|^p w(t)dt\,. \end{aligned}$$

We also need the following lemma.

Lemma 3 *[Csörgő and Horváth (1988; Lemma 7).]*
Let $X_1, \cdots, X_n$ be iid random variables with the cdf F. Let $\alpha_n(t) = \sqrt{n}(F_n(t) - F(t))$ be the corresponding empirical process, where $F_n(t) = n^{-1}\sum_{i=1}^n I\{X_i \le t\}$. Then we can find a sequence of Brownian bridges $\{B_n(t),\, 0 \le t \le 1\}$, $n = 1, 2, \cdots$, such that

$$\sup_{-\infty<t<\infty} \frac{|\alpha_n(t) - B_n(F(t))|}{[F(t)(1-F(t))]^{\frac{1}{2}-\nu}} = \mathcal{O}_p(\kappa_\nu(n))\,,$$

where κ_ν is as in (6).

By Lemma 3, the are two independent sequences of Brownian bridges $\{B_n^{(1)}(t)\,, 0 \le t \le 1\}_{n\ge1}$ and $\{B_m^{(2)}(t)\,, 0 \le t \le 1\}_{m\ge1}$ such that

$$\sup_{t\in\mathbb{R}} \frac{|\alpha_n^{(1)}(t) - B_n^{(1)}(F_1(t))|}{[F_1(t)(1-F_1(t))]^{\frac{1}{2}-\nu}} = \mathcal{O}_p(\kappa_\nu(n)) \tag{33}$$

$$\text{and} \quad \sup_{t\in\mathbb{R}} \frac{|\alpha_m^{(2)}(t) - B_m^{(2)}(F_2(t))|}{[F_2(t)(1-F_2(t))]^{\frac{1}{2}-\nu}} = \mathcal{O}_p(\kappa_\nu(m)) \tag{34}$$

Let $B_n^{(1)}(\cdot)$ and $B_m^{(2)}(\cdot)$ be as in (33) and (34) and observe that when $f_1 = f_2$ $(= f)$, we have

$$\begin{aligned} &\left(\frac{m+n}{mn}\right)^{-\frac{p}{2}} h^p \cdot I_{m,n}(p) = \\ &\int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x))\right\} \right|^p w(t)dt + R_{m,n} \end{aligned} \tag{35}$$

where

$$\begin{aligned} R_{m,n} &= \int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} \alpha_n^{(1)}(x) - \sqrt{\frac{n}{m+n}} \alpha_m^{(2)}(x)\right\} \right|^p w(t)dt \\ &\quad - \int \left| \int K\left(\frac{t-x}{h}\right) d\left\{\sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x))\right\} \right|^p w(t)dt \end{aligned}$$

Using the inequality $||a|^p - |b|^p| \le p2^{p-1}|a-b|^p + p2^{p-1}|b|^{p-1}|a-b|$, (for $p \ge 1$), one finds

$$
\begin{aligned}
&\frac{1}{p2^{p-1}}|R_{m,n}| \\
&\leq \Bigg[\int\Bigg|\int K\left(\frac{t-x}{h}\right)d\Bigg\{\sqrt{\frac{m}{m+n}}\alpha_n^{(1)}(x)-\sqrt{\frac{n}{m+n}}\alpha_m^{(2)}(x) \\
&\qquad\qquad -\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))+\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\Bigg\}\Bigg|^p w(t)dt\Bigg] \\
&\quad+\Bigg[\int\Bigg|\int K\left(\frac{t-x}{h}\right)d\Bigg\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))+\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\Bigg\}\Bigg|^{p-1} \\
&\qquad\times\Bigg|\int K\left(\frac{t-x}{h}\right)d\Bigg\{\sqrt{\frac{m}{m+n}}\alpha_n^{(1)}(x)-\sqrt{\frac{n}{m+n}}\alpha_m^{(2)}(x) \\
&\qquad\qquad -\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))+\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\Bigg\}\Bigg|w(t)dt\Bigg] \\
&\leq \Bigg[\int\Bigg|\int K\left(\frac{t-x}{h}\right)d\Bigg\{\sqrt{\frac{m}{m+n}}\alpha_n^{(1)}(x)-\sqrt{\frac{n}{m+n}}\alpha_m^{(2)}(x) \\
&\qquad\qquad -\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))+\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\Bigg\}\Bigg|^p w(t)dt\Bigg] \\
&\quad+\Bigg[\left(\int\Bigg|\int K\left(\frac{t-x}{h}\right)d\Bigg\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))+\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\Bigg\}\Bigg|^p w(t)dt\right)^{\frac{p-1}{p}} \\
&\qquad\times\left(\int\Bigg|\int K\left(\frac{t-x}{h}\right)d\Bigg\{\sqrt{\frac{m}{m+n}}\left(\alpha_n^{(1)}(x)-B_n^{(1)}(F(x))\right)\right. \\
&\qquad\qquad\left.-\sqrt{\frac{n}{m+n}}\left(\alpha_m^{(2)}(x)-B_m^{(2)}(F(x))\right)\Bigg\}\Bigg|^p w(t)dt\right)^{\frac{1}{p}}\Bigg] \\
&\qquad\text{(where the last large square-bracketed term follows by Hölder's inequality)}, \\
&:= \left[R_{m,n}(I)\right]+\left[R_{m,n}(II)\right] \qquad (36)
\end{aligned}
$$

However,

$$
\begin{aligned}
R_{m,n}(I) &\leq 2^{p-1}\int\left|\int K\left(\frac{t-x}{h}\right)\sqrt{\frac{m}{m+n}}d\left\{\alpha_n^{(1)}(x)-B_n^{(1)}(F(x))\right\}\right|^p w(t)dt \\
&\quad+2^{p-1}\int\left|\int K\left(\frac{t-x}{h}\right)\sqrt{\frac{n}{m+n}}d\left\{\alpha_n^{(2)}(x)-B_m^{(2)}(F(x))\right\}\right|^p w(t)dt \\
&:= 2^{p-1}\Big(R_{m,n,1}(I)+R_{m,n,2}(I)\Big), \quad \text{(say)}. \qquad (37)
\end{aligned}
$$

Now, let $a>0$ be a constant such that $K(t)=0$ for $t\notin(-a,a)$, (see condition $\mathcal{K}(iii)$) and

observe that

$$
\begin{aligned}
& \left| \int K\left(\frac{t-x}{h}\right) \sqrt{\frac{m}{m+n}}\, d\left\{\alpha_n^{(1)}(x) - B_n^{(1)}(F(x))\right\} \right|^p \\
= & \left(\frac{m}{m+n}\right)^{\frac{p}{2}} \left| \int_{-a}^{a} \frac{\alpha_n^{(1)}(t-yh) - B_n^{(1)}(F(t-yh))}{[F(t-yh)(1-F(t-yh))]^{0.5-\nu}} [F(t-yh)(1-F(t-yh))]^{0.5-\nu} dK(y) \right|^p \\
\leq & \left(\frac{m}{m+n}\right)^{\frac{p}{2}} \left(\sup_{s\in\mathbb{R}} \frac{\left|\alpha_n^{(1)}(s) - B_n^{(1)}(F(s))\right|}{[F(s)(1-F(s))]^{0.5-\nu}} \times \sup_{-a\leq y\leq a} [F(t-yh)(1-F(t-yh))]^{0.5-\nu} \int_{-a}^{a} dK(y) \right)^p \\
= & \mathcal{O}(1) \cdot \mathcal{O}_p(\kappa_\nu^p(n)) \cdot [F(t+ah)(1-F(t-ah))]^{0.5-\nu} \qquad (38) \\
= & \mathcal{O}_p(\kappa_\nu^p(n)) \cdot [F(t+ah)(1-F(t-ah))]^{0.5-\nu},
\end{aligned}
$$

where in (38) the $\mathcal{O}(1)$ term follows from the fact that $\frac{m}{m+n} = \mathcal{O}(1)$, whereas the $\mathcal{O}_p(\kappa_\nu^p(n))$ term follows from Lemma 3. The last term in (38) follows because $p \geq 1$ and $0 \leq F(t+ah)(1-F(t-ah)) \leq 1$. Since the $\mathcal{O}_p(\kappa_\nu^P(n))$ term in (38) does not depend on t, we find

$$
\begin{aligned}
R_{m,n,1}(I) & \leq \mathcal{O}_p(\kappa_\nu^p(n)) \cdot \int [F(t+ah)(1-F(t-ah))]^{0.5-\nu} w(t)dt \\
& = \mathcal{O}_p(\kappa_\nu^p(n)), \quad (\text{by } \mathcal{F}(iv)).
\end{aligned}
$$

Similarly, we have $R_{m,n,2}(I) = \mathcal{O}_p(\kappa_\nu^p(m))$. Therefore,

$$
R_{m,n}(I) = \mathcal{O}_p(\kappa_\nu^p(m \wedge n)). \qquad (39)
$$

To deal with the term $R_{m,n}(II)$ in (36), first observe that

$$
\begin{aligned}
R_{m,n}(II) = & \left(\int \left| \int K\left(\frac{t-x}{h}\right) d\left\{ \sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x)) \right\} \right|^p w(t)dt \right)^{\frac{p-1}{p}} \\
& \times (R_{m,n}(I))^{\frac{1}{p}}. \qquad (40)
\end{aligned}
$$

Now, let $\{B^{(1)}(t), 0 \leq t \leq 1\}$ and $\{B^{(2)}(t), 0 \leq t \leq 1\}$ be independent Brownian bridges and note that for each $m = 1, 2, \cdots$, and $n = 1, 2, \cdots$,

$$
\begin{aligned}
& \left\{ \int K\left(\frac{t-x}{h}\right) d\left\{ \sqrt{\frac{m}{m+n}} B_n^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B_m^{(2)}(F(x)) \right\},\ t \in \mathbb{R} \right\} \\
& \quad \stackrel{\mathrm{d}}{=} \left\{ \int K\left(\frac{t-x}{h}\right) d\left\{ \sqrt{\frac{m}{m+n}} B^{(1)}(F(x)) - \sqrt{\frac{n}{m+n}} B^{(2)}(F(x)) \right\},\ t \in \mathbb{R} \right\}. \quad (41)
\end{aligned}
$$

Next, let $B(\cdot)$ be a Brownian bridge on $[0,1]$ and note that by the standard properties of independent Brownian bridges (see, for example, Shorack and Wellner (1986; page 32)),

$$\left\{\sqrt{\frac{m}{m+n}}B^{(1)}(s)-\sqrt{\frac{n}{m+n}}B^{(2)}(s)\,,\ s\in[0,1]\right\} \stackrel{\mathrm{d}}{=} \left\{B(s)\,,\ s\in[0,1]\right\}. \tag{42}$$

Furthermore, by Lemma 2

$$\int|\Gamma_n(t)|^p\,w(t)dt \;=\; \mathcal{O}_p(h^{\frac{p}{2}})\,,\ \text{where } \Gamma_n(t)=\int K\left(\frac{t-x}{h}\right)d(W(F(x))-F(x)W(1))\,.$$

But, $\{W(s)-sW(1)\,,0\le s\le 1\}\stackrel{\mathrm{d}}{=}\{B(s)\,,0\le s\le 1\}$, where $W(\cdot)$ is a standard Brownian motion. Therefore, in view of (41) and (42),

$$\int\left|\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))-\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\}\right|^p w(t)dt \;=\; \mathcal{O}_p(h^{\frac{p}{2}})\,.$$

Combining this fact with (39) and (40), we find

$$R_{m,n}(II) \;=\; \mathcal{O}_p(h^{\frac{p-1}{2}})\mathcal{O}_p(\kappa_\nu(m\wedge n)) \;=\; \mathcal{O}_p(h^{\frac{p-1}{2}}\kappa_\nu(m\wedge n)). \tag{43}$$

Putting together (43), (39), and (36), we find

$$|R_{m,n}| \;=\; \mathcal{O}_p(\kappa_\nu^p(m\wedge n))+\mathcal{O}_p(h^{\frac{p-1}{2}}\kappa_\nu(m\wedge n)). \tag{44}$$

Next, observe that

$$\begin{aligned}
&\left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}\left\{\left(\frac{m+n}{mn}\right)^{-\frac{p}{2}}h^p\cdot I_{m,n}(p)-h^{\frac{p}{2}}\eta\right\}\\
&\stackrel{\text{by (35)}}{=}\left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}\left(\int\left|\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))\right.\right.\right.\\
&\qquad\left.\left.\left.-\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\}\right|^p w(t)dt-h^{\frac{p}{2}}\eta\right)+\left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}R_{m,n}\,.
\end{aligned}$$

The independence of the Brownian bridges $B_n^{(1)}(\cdot)$ and $B_m^{(2)}(\cdot)$, (for all $m=1,2,\cdots$ and $n=1,2,\cdots$), in conjunction with (41) and Lemma 2 yields

$$\begin{aligned}
&\left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}\left(\int\left|\int K\left(\frac{t-x}{h}\right)d\left\{\sqrt{\frac{m}{m+n}}B_n^{(1)}(F(x))\right.\right.\right.\\
&\qquad\left.\left.\left.-\sqrt{\frac{n}{m+n}}B_m^{(2)}(F(x))\right\}\right|^p w(t)dt-h^{\frac{p}{2}}\eta\right)\xrightarrow{\mathrm{d}}N(0,1)\,.
\end{aligned}$$

Furthermore,

$$\begin{aligned}\left(h^{p+1}\sigma^2\right)^{-\frac{1}{2}}|R_{m,n}| &\overset{\text{by (44)}}{=} h^{-\frac{p+1}{2}}\kappa_\nu^p(m\wedge n)\cdot\mathcal{O}_p(1)+h^{-\frac{p+1}{2}+\frac{p-1}{2}}\cdot\kappa_\nu(m\wedge n)\cdot\mathcal{O}_p(1)\\ &= o_p(1)+o_p(1)\,,\end{aligned}$$

where the last line follows from the condition $h^{-1}\cdot\kappa_\nu(m\wedge n)\to 0$ and the fact that $h^{-\frac{p+1}{2}}\leq h^{-p}$, for all $p\geq 1$. The theorem now follows from the above results together with Slutsky's theorem.

□

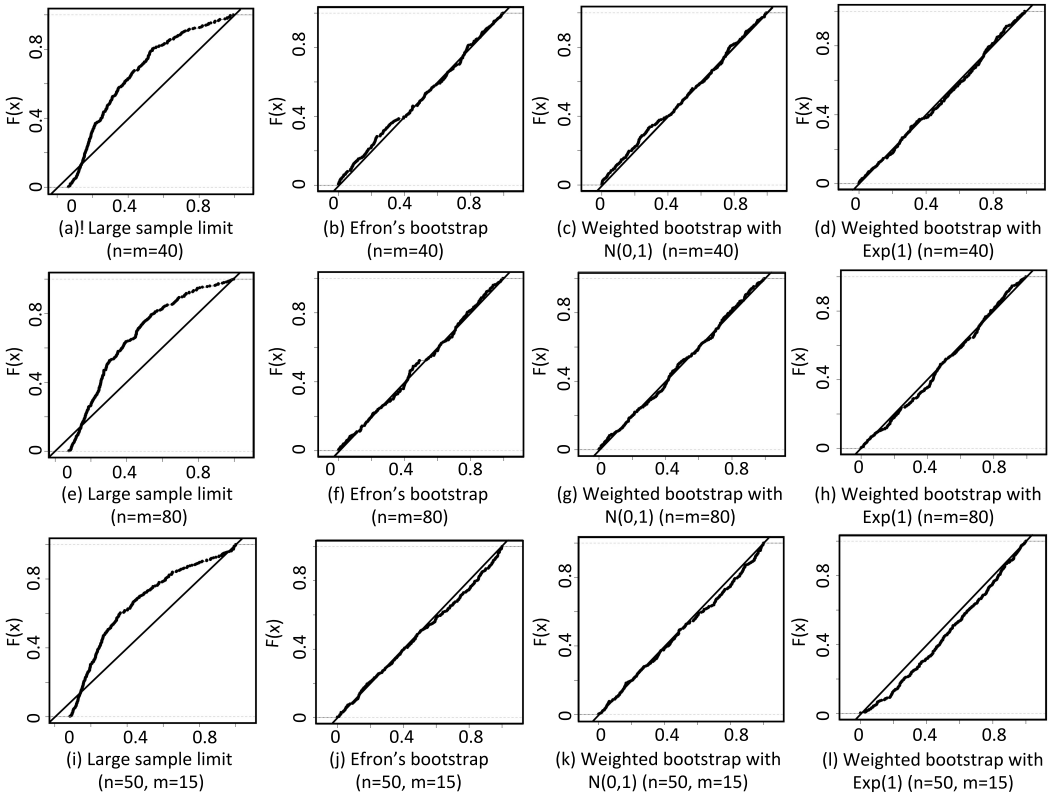

Figure 1: Plots of empirical cdf's when h=0.25: (a), (e), (i) correspond to $U_1, \cdots, U_{400}$, (b), (f), (j) correspond to $V_1, \cdots, V_{400}$, and (c), (d), (g), (h), (k), (l) correspond to $Y_1, \cdots, Y_{400}$.

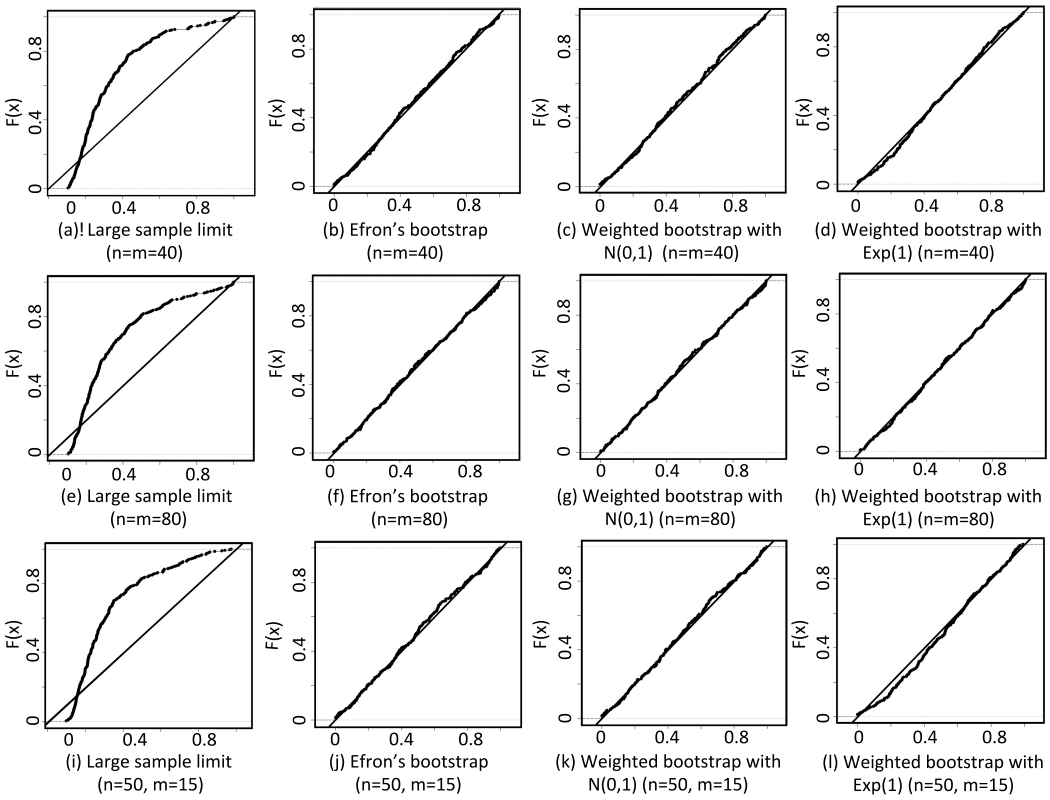

Figure 2: Plots of empirical cdf's when h=0.40: (a), (e), (i) correspond to $U_1, \cdots, U_{400}$, (b), (f), (j) correspond to $V_1, \cdots, V_{400}$, and (c), (d), (g), (h), (k), (l) correspond to $Y_1, \cdots, Y_{400}$.

Acknowledgements

The authors would like to thank Editor, Iréne Gijbels, Associate Editor, and anonymous referees for their comments on the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by the National Science Foundation under 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', *Statistical Methodology*, 11, 36-52.

Anderson, H., Hall, P., and Titterington, D. M. (1994), 'Two-Sample Test Statistics for Measuring Discrepancies Between two Multivariate Probability Density Functions Using Kernel-based Density Estimates', *Journal of Multivariate Analysis*, 50, 41–54.

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

Beirlant, J., and Mason, D. (1995), 'On the Asymptotic Normality of L_p-norms of Empirical Functionals', *Mathematical Methods of Statistics*, 4, 1-19.

Berlinet, A., Devroye, L., and Gyorfi, 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' *Annals of Statistics*, 1, 1075-1095.

Bowman, A. (1984), 'An Alternative Method of Cross-Validation for the Smoothing of Density Estimates', *Biometrika*, 71, 353-360.

Burke, M. (1998), 'A Gaussian Bootstrap Approach to Estimation and Tests' in *Asymptotic Methods in Probability and Statistics*, ed. B. Szyszkowicz, Amsterdam: North-Holland, 697-706.

Burke, M. (2000), 'Multivariate Tests-of-fit and Uniform Confidence Bands Using a Weighted Bootstrap', *Statistics and Probability Letters*, 46, 13-20.

Burke, M. (2010), 'Approximations for a Multivariate Hybrid Process with Applications to Change-point Detection', *Mathematical Methods of Statistics*, 19, 121-135.

Cao, R., and Van Keilegom, I. (2006), 'Empirical Likelihood Tests for Two-sample Problems via Nonparametric Density Estimation.' *Canadian Journal of Statistics*, 34, 61–77.

Csörgő, M., and Horváth, L. (1988), 'Central Limit Theorems for L_p-norms of Density Estimators', *Probability Theory and Related Fields*, 80, 269-291.

Csörgő, M., and Horváth, L. (1993), *Weighted Approximations in Probability and Statistics*, Chichester: Wiley.

Efron, B. (1979), 'Bootstrap Methods: Another Look at the Jackknife', *Annals of Statistics*, 7, 1-26.

Eggermont, P., and LaRiccia, V. N. (2001), *Maximum Penalized Likelihood Estimation Vol. I. Density estimation*, New York: Springer-Verlag.

Fryer, M.J. (1977), 'A Review of Some Non-parametric Methods of Density Estimation', *Journal of the Institute of Mathematics and its Applications*, 20, 335-354.

Hall, P. (1982), 'Limit Theorems for Stochastic Measures of the Accuracy of Density Estimators', *Stochastic Processes and their Applications*, 13, 11-25.

Hall, P. (1984), 'Central Limit Theorems for Integrated Square Error of Multivariate Nonparametric Density Estimators', *Journal of Multivariate Analysis*, 14, 1-16.

Hall, P., and Mammen, E. (1994), 'On General Resampling Algorithms and Their Perfor-

mance in Distribution Estimation', *Annals of Statistics*, 22, 2011-2030.

Henze, N., and Nikitin, Y. (2003), 'Two-Sample Tests Based on the Integrated Empirical Process', *Communications in Statistics: Theory Methods*, 32, 1767–1788.

Henze, N., Nikitin, Y., and Ebner, B. (2009), 'Integral Distribution-Free Statistics of L_p-type and Their Asymptotic Comparison', *Computational Statistics and Data Analysis*, 53, 3426–3438.

Horváth, L. (1991), 'On L_p-norms of Multivariate Density Estimators', *Annals of Statistics*, 19, 1933-1949.

Horváth, L., Kokoszka, P., and Steinebach, J. (2000), 'Approximations for Weighted Bootstrap Processes with an Application', *Statistics and Probability Letters*, 48, 59-70.

Kojadinovix, I., Yan, J. (2011). 'Goodness-if-fit Testing BAsaed on a Weighted Boostrap: A Fast Large-sample Alternative to the Parameteric Boostrap', *The Canadian Journal of Statistics*, 40, 480-500.

Kojadinovic, I., Yan, J., and Holmes, M. (2011). 'Fast Large-sample Goodness-of-fit for Copulus', *Statistica Sinica*, 221, 841-871

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

Liu, B. and Mojirsheibani, M. (2015). 'On a Weighted Bootstrap Approximation of L_p Norms of Kernel Density Estimators', *Statistics and Probability Letters*, 105, 65-73

Mason, D., and Newton, M. (1992), 'A Rank Statistics Approach to the Consistency of a General Bootstrap', *Annals of Statistics*, 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', *Annals of Mathematical Statistics*, 33, 1065-1076.

Prakasa Rao, B. L. S. (1983), *Nonparametric Functional Estimation*, Orlando: Academic Press.

Rosenblatt, M. (1952), 'Limit Theorems Associated with Variants of the von Mises Statistic',

Annals of Mathematical Statistics, 23, 617–623.

Rosenblatt, M. (1956), 'Remarks on Some Nonparametric Estimates of Density Functions', *Annals of Mathematical Statistics*, 27, 832-837.

Rubin, D. (1981), 'The Bayesian bootstrap', *Annals of Statistics*, 9, 130-134.

Rudemo, M. (1982), 'Empirical Choice of Histograms and Kernel Density Estimators', *Scandinavian Journal of Statistics*, 9, 65-78.

Sheather, S. and Jones, M. (1991), 'A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation', *Journal of the Royal Statistical Society: Ser. B*, 53, 683-690.

Shorack, G., and Wellner, J. (1986), *Empirical Processes with Applications to Statistics*, New York: John Wiley & Sons.

Steele, J. M. (1978), 'Invalidity of Average Squared Error Criterion in Density Estimation', *The Canadian Journal of Statistics*, 6, 193-200.

Wand, M. and Jones, M. (1994), 'Multivariate Plug-in Bandwidth Selection', *Computational Statistics*, 9, 97-116.

Wegman, E.J. (1972), 'Nonparametric Probability Density Estimation: A Comparison of Density Estimation Methods', *Journal of Statistical Computation and Simulation*, 1, 225-245.