Journal of Statistical Research *ISSN 0256 - 422 X*
2010, Vol. 44, No. 2, pp. 219-232
Bangladesh

A NOTE ON THE WEIGHTED BOOTSTRAP APPROXIMATION OF THE BICKEL-ROSENBLATT STATISTIC

Majid Mojirsheibani
School of Mathematics and Statistics, Carleton University,
Ottawa, Ontario, K1S 5B6, Canada.
Email: majidm@math.carleton.ca

SUMMARY

In this article we propose a weighted bootstrap approximation to the distribution of the supremum (over compact sets) of the classical Bickel-Rosenblatt statistic $|f_n(t) - f(t)|/\sqrt{f(t)}$ as well as its "Studentized" version $|f_n(t) - f(t)|/\sqrt{f_n(t)}$, where f_n is the usual kernel density estimator of the true density f. Following Horvath et al. (2000), we showed that the proposed weighted bootstrap method is consistent (in capturing the true limiting distributions derived by Bickel and Rosenblatt (1973)). Furthermore, simulation results show that the proposed weighted bootstrap has a much better finite-sample performance than the results based on asymptotic theory. For comparison purposes, we also consider Efron's (1979) original bootstrap.

Keywords and phrases: Place keywords here

AMS Classification: Place Classification here. Leave as is, if there is no classification

1 Introduction

Let $X_1, \cdots, X_n$ be independent and identically distributed random variables with a common distribution function F and the probability density function $f = F'$. Also, let

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

be the usual Parzen-Rosenblatt kernel density estimate of f (Parzen (1962) and Rosenblatt (1956)), where K, the kernel function, is typically required to satisfy certain regularity conditions.; here h_n is the smoothing parameter of the kernel. An important statistic, which is also a standard measure of the performance of f_n, as an estimator of f, is given by the function $\sup_{a\leq t\leq b} |f_n(t)-f(t)|/\sqrt{f(t)}$, where $-\infty < a < b < \infty$. Using the theory of the extrema of Gaussian Processes, Bickel and Rosenblatt (1973) derived the asymptotic distribution of the above statistic as well as its "Studentizes" version, $\sup_{a\le t\le b}|f_n(t)-f(t)|/\sqrt{f_n(t)}$, under appropriate regularity conditions. More specifically, suppose that the following conditions hold:

Condition (K)

K**(i).** The kernel K is nonnegative, symmetric about zero, and vanishes outside an interval $[-A, A]$, for some $A < \infty$.

K**(ii).** K is uniformly bounded and $\int K(x)\,d(x) = 1$. The derivative K' exists (a.e.) on $(-A, A)$ and satisfies $\int x^2\,|K'(x)|\,dx < \infty$.

K**(iii).** $K(x) = L(|x|^p)$ for some nonincreasing function L, where $p \ge 1$ is an integer.

An example of a kernel that satisfies all the above conditions is a truncated Gaussian kernel (truncated at $-A$ and A).

Condition (f)

The density f is continuous, bounded, and positive on $(-\epsilon, 1+\epsilon)$, for some $\epsilon > 0$. Furthermore, the function $f^{1/2}$ is absolutely continuous and its derivative is bounded in absolute value. Also, f'' exists and is bounded.

Let

$$M_n = \sup_{0\le t\le 1}\sqrt{\frac{nh_n}{f(t)}}\,\Big|f_n(t)-f(t)\Big| \qquad \text{and} \qquad \widehat{M}_n = \sup_{0\le t\le 1}\sqrt{\frac{nh_n}{f_n(t)}}\,\Big|f_n(t)-f(t)\Big|.$$

Although the supremum is taken over the interval $[0,1]$, the results hold on any other interval on which f is bounded away fron 0 and ∞. Then, Bickel and Rosenblatt (1973) proved the following results.

Theorem 1. *Suppose that conditions* K(i)*,* K(ii)*, and* (f) *hold. If*

$$h_n = n^{-\delta}, \quad \textit{where}\ \ \frac{1}{5} < \delta < \frac{1}{2}$$

then, as $n \to \infty$*,*

$$\begin{aligned}\lim_{n\to\infty} P\left\{\sqrt{2\delta\,\log n}\left(\frac{M_n}{\lambda^{1/2}} - d_n\right) \le x\right\} &= \lim_{n\to\infty} P\left\{\sqrt{2\delta\,\log n}\left(\frac{\widehat{M}_n}{\lambda^{1/2}} - d_n\right) \le x\right\} \\ &= \exp\left(-2e^{-x}\right), \qquad (1.1)\end{aligned}$$

where $\lambda = \int K^2(u)du$,

$$d_n = \sqrt{2\delta \log n} + \begin{cases} \frac{\log K_1 - \frac{1}{2} \log \pi + \frac{1}{2}(\log \delta + \log \log n)}{(2\delta \log n)^{1/2}}, & \text{if } K_1 := \frac{K^2(A)+K^2(-A)}{2\lambda} > 0, \\ \frac{\log\left[(1/\pi)(K_2/2)^{1/2}\right]}{(2\delta \log n)^{1/2}}, & \text{otherwise}, \end{cases} \tag{1.2}$$

and where $K_2 = \frac{1}{2\lambda} \int (K'(t))^2 dt$.

From a statistical point of view, the result in (1.1) can be used to form asymptotic confidence bands for the unknown density f. However, it is also well known that the rate of convergence in (1.1) is very slow; see, for example, Konakov and Piterbarg (1984) and Hall (1991). An alternative approach to approximate the limiting distribution in (1.1) is based on the bootstrap. Efron's (1979) original bootstrap algorithm replaces M_n and $\widehat{M}_n$ in (1.1) by M_n^* and $\widehat{M}_n^*$, respectively, where

$$M_n^* = \sup_{0 \le t \le 1} \sqrt{\frac{nh_n}{f_n(t)}} \left| f_n^*(t) - f_n(t) \right| \quad \text{and} \quad \widehat{M}_n^* = \sup_{0 \le t \le 1} \sqrt{\frac{nh_n}{f_n^*(t)}} \left| f_n^*(t) - f_n(t) \right| ;$$

here

$$f_n^*(t) = (nh_n)^{-1} \sum_{i=1}^{n} K((t - X_i^*)/h_n) \tag{1.3}$$

where $X_1^*, \cdots, X_n^*$ are conditionally independent (conditional on $X_1, \cdots, X_n$) with distribution function $F_n(t) = n^{-1} \sum_{i=1}^n I\{X_i \le t\}$. The theoretical validity of the corresponding bootstrap versions of (1.1) follows from the work of Csörgő et al. (2000) and Hall (1991). For $1 \le p < \infty$, one may also consider L_p-norms of kernel density estimators (and their bootstrap versions) for goodness-of-fit tests; see, for example, Horvath (1991) and Mojirsheibani (2007).

In this article we consider a weighted bootstrap approach for approximating the limiting distributions in (1.1). The proposed approach is very easy to implement and produces accurate approximations.

2 Main Results

Many authors have used the weighted bootstrap as a generalization of Efron's (1979) original bootstrap in the literature. Burke (2000) uses Gaussian weights to form bootstrap confidence bands for a distribution functions. Mason and Newton (1992) give conditions under which the weighted bootstrapped mean is consistent. Horvath 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. Another interesting application of the weighted bootstrap is the Bayesian bootstrap of Rubin (1981). One may also refer to Barbe and Bertail (1995) for a general view and further results on the weighted bootstrap. Of course, Efron's original bootstrap itself is a weighted bootstrap algorithm, where the weights are multinomial 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), and Horvath et al. (2000).

The weighted bootstrap approximation of this article works as follows. Let $\epsilon_1, \cdots, \epsilon_n$ be i.i.d. random variables, with mean μ and variance 1, independent of the data $X_1, \cdots, X_n$. The weighted bootstrap version of f_n is given by

$$f_{nn}(t) = (nh_n)^{-1} \sum_{i=1}^{n} (1 + \epsilon_i - \bar{\epsilon}) K((t - X_i)/h_n), \tag{2.1}$$

where $\bar{\epsilon} = n^{-1} \sum_{i=1}^{n} \epsilon_i$. Observe that if we replace $(1 + \epsilon_i - \bar{\epsilon})$ by $M_{n,i}$ in (2.1), where $(M_{n,1}, \cdots, M_{n,n})$ is a multinomial random vector with n draws on n categories, then $f_{nn}(t)$ reduces to $f_n^*(t)$ in (1.3). Next, consider the following counterparts of M_n and $\widehat{M}_n$:

$$M_{nn} = \sup_{0 \le t \le 1} \sqrt{\frac{nh_n}{f_n(t)}} \left| f_{nn}(t) - f_n(t) \right|$$

and

$$\widehat{M}_{nn} = \sup_{0 \le t \le 1} \sqrt{\frac{nh_n}{f_{nn}(t)}} \left| f_{nn}(t) - f_n(t) \right|.$$

The following theorem shows that our weighted bootstrap approximation of (1.1) works correctly. We first need to state a condition regarding the random variables $\epsilon_1, \cdots, \epsilon_n$ used in (2.1).

Condition (M)

The random variables $\epsilon_1, \cdots, \epsilon_n$ are i.i.d., with some mean μ and variance 1, and are independent of the data $X_1, \cdots, X_n$. Furthermore, there is a $t_0 > 0$ such that $E(e^{t\epsilon_1}) < \infty$ for all $t \in (-t_0, t_0)$.

Theorem 2. *Suppose that conditions (K) , (f), and (M) hold. If*

$$h_n = n^{-\delta}, \quad \text{where } \frac{1}{5} < \delta < \frac{1}{2}$$

then, as $n \to \infty$,

$$\begin{aligned} \lim_{n\to\infty} P\left\{ \sqrt{2\delta \log n} \left(\frac{M_{nn}}{\lambda^{1/2}} - d_n \right) \le x \right\} &= \lim_{n\to\infty} P\left\{ \sqrt{2\delta \log n} \left(\frac{\widehat{M}_{nn}}{\lambda^{1/2}} - d_n \right) \le x \right\} \\ &= \exp\left(-2e^{-x}\right), \end{aligned} \tag{2.2}$$

where $\lambda = \int K^2(u)du$ and d_n is as in (1.2).

Proof. Let

$$\begin{aligned} F_n(t) &= n^{-1}\sum_{i=1}^{n} I\{X_i \le t\} \\ F_{nn}(t) &= n^{-1}\sum_{i=1}^{n}(1+\epsilon_i-\bar{\epsilon})I\{X_i \le t\} \\ \beta_n(t) &= n^{1/2}(F_{nn}(t)-F_n(t)) \;=\; n^{-1/2}\sum_{i=1}^{n}(\epsilon_i-\bar{\epsilon})I\{X_i \le t\} \end{aligned}$$

and observe that

$$\begin{aligned} f_{nn}(t)-f_n(t) &= (nh_n)^{-1}\left[\sum_{i=1}^{n}(1+\epsilon_i-\bar{\epsilon})K((t-X_i)/h_n)-\sum_{i=1}^{n}K((t-X_i)/h_n)\right] \\ &= \frac{1}{h}\int_{-\infty}^{\infty}K((t-s)/h_n)\,d(F_{nn}(s)-F_n(s)). \end{aligned}$$

Therefore

$$(nh_n)^{1/2}(f_{nn}(t)-f_n(t)) = h^{-1/2}\int_{-\infty}^{\infty}K((t-s)/h_n)\,d\beta_n(s).$$

Next, we need to state the following result of Horvath et al. (2000):

Lemma 2.1. *If the weights $\epsilon_1,\cdots,\epsilon_n$ satisfy condition (M) then there exists a sequence of Brownian bridges $\{B_n(t),\ 0\le t\le 1\}$ such that*

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

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

The following corollary is an immediate consequence of Lemma 2.1.

Corollary 2.1. Under condition (M),

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

where $\{B_n(t),\ 0\le t\le 1\}$ is as in Lemma 2.1.

Now, let $\{B_n(t),\ 0\le t\le 1\}$ be as in Lemma 2.1 and observe that

$$\begin{aligned} &\sup_{0\le t\le 1}\left|\sqrt{\frac{nh_n}{f(t)}}\left(f_{nn}(t)-f_n(t)\right)-\frac{1}{\sqrt{h\,f(t)}}\int_{-\infty}^{\infty}K((t-s)/h_n)\,dB_n(F(s))\right| \\ =\;& h_n^{-1/2}\sup_{0\le t\le 1}\frac{1}{\sqrt{f(t)}}\left|\int_{-\infty}^{\infty}\beta_n(t-xh_n)\,dK(x)-\int_{-\infty}^{\infty}B_n(F(t-xh))\,dK(x)\right| \\ \le\;& h_n^{-1/2}\left|\int_{-\infty}^{\infty}d\,K(x)\right|\sup_{0\le t\le 1}\frac{1}{\sqrt{f(t)}}\times\sup_{-\infty<u<\infty}\left|\beta_n(u)-B_n(F(u))\right| \\ =\;& O_P\left(\frac{\log n}{\sqrt{nh_n}}\right),\quad \text{(by Corollary 2.1 and the assumptions on } f \text{ and } K). \qquad (2.3) \end{aligned}$$

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

$$\left\{ \frac{1}{\sqrt{h_n f(t)}} \int_{-\infty}^{\infty} K((t-s)/h_n)\, dB_n(F(s)),\ 0 \le t \le 1 \right\}$$
$$\stackrel{d}{=} \left\{ \frac{1}{\sqrt{h_n f(t)}} \int_{-\infty}^{\infty} K((t-s)/h_n)\, dB(F(s)),\ 0 \le t \le 1 \right\}.$$

Bickel and Rosenblatt (1973) studied the process $(h_n f(t))^{-1/2} \int_{-\infty}^{\infty} K((t-s)/h_n) d\, B(F(s))$, $0 \le t \le 1$, and showed that its normalized supremum, i.e.,

$$\sqrt{2\delta \, \log n} \left(\lambda^{-1/2} \sup_{0\le t\le 1} \left| \frac{1}{\sqrt{h_n f(t)}} \int_{-\infty}^{\infty} K((t-s)/h_n)\, dB(F(s)) \right| - d_n \right),$$

converges in distribution to a random variable Y, where $P\{Y \le x\} = \exp(-2e^{-x})$. Thus, in view of (2.3),

$$P\left\{ \sqrt{2\delta \, \log n} \left(\lambda^{-1/2} \sup_{0\le t\le 1} \left| \sqrt{\frac{nh_n}{f(t)}} \Big(f_{nn}(t) - f_n(t) \Big) \right| - d_n \right) \le x \right\} = \exp\left(-2e^{-x}\right). \quad (2.4)$$

Next, we show that

$$\sqrt{(nh_n)\log n} \sup_{0\le t\le 1} \left| \left(\frac{1}{\sqrt{f_n(t)}} - \frac{1}{\sqrt{f(t)}} \right) \big(f_{nn}(t) - f_n(t) \big) \right| = o_P(1). \quad (2.5)$$

We observe from (2.5) and (2.4) that the first limit statement of Theorem 2 is equal to $\exp(-2e^{-x})$. To establish (2.5), first note by (2.4) that we have

$$\sup_{0\le t\le 1} \left| \sqrt{nh_n/f(t)}\, (f_{nn}(t) - f_n(t)) \right| = O_P\left(\sqrt{\log n} \right).$$

Therefore, the left side of (2.5) is bounded by

$$\begin{aligned} \sup_{0\le t\le 1} \left| \frac{\sqrt{f(t)}}{\sqrt{f_n(t)}} - 1 \right| \times O_P(\log n) \quad &\le \quad \left(\frac{1}{\inf_{0\le t\le 1} f_n(t)} \right)^{1/2} \sup_{0\le t\le 1} \frac{|f_n(t) - f(t)|}{\sqrt{f(t)}} \times O_P(\log n) \\ &= \quad O_P(1) \times O_P\left(\sqrt{\frac{\log n}{nh_n}} \right) \times O_P(\log n) \ = \ o_P(1), \end{aligned}$$

where the $O_P(1)$ term follows from a result of Devroye (1978; eq. (21)), whereas the $O_P\big(\sqrt{\log n/(nh_n)}\,\big)$ term is a direct consequence of the first limit statement in (1.1). This completes the proof of (2.5). Similarly, the second limit statement in Theorem 2 follows from (2.4) and the observation that

$$\sqrt{(nh_n)\log n} \sup_{0\le t\le 1} \left| \left(\frac{1}{\sqrt{f_{nn}(t)}} - \frac{1}{\sqrt{f(t)}} \right) \big(f_{nn}(t) - f_n(t) \big) \right| = o_P(1).$$

$\square$

2.1 Numerical Examples

To illustrate the theoretical findings of the paper, we provided a numerical assessment of the performance and effectiveness of the methods discussed in this section. It will be seen that, in general, the weighted bootstrap performs very well in capturing the finite-sample distribution of the Bickel-Rosenblatt statistic. Our examples involve random samples of sizes $n = 50$ and $n = 100$ drawn from the mixture of normals:

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

As for the choice of the kernel, we considered the truncated Gaussin density

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

where Φ is the N(0, 1) distribution function. Next we computed the kernel density estimate f_n for each of the samples as well as their corresponding Bickel-Rosenblatt statistics,

$$Y_n := \sqrt{2\delta\,\log n}\,\left(\frac{M_n}{\lambda^{1/2}} - d_n\right) \quad \text{and} \quad \widehat{Y}_n := \sqrt{2\delta\,\log n}\,\left(\frac{\widehat{M}_n}{\lambda^{1/2}} - d_n\right)$$

for 7 different values of the smoothing parameter $h_n = 0.10,\ 0.15,\ 0.20,\ 0.25,\ 0.30,\ 0.35,\ 0.40$, (recall that $h_n = n^{-\delta}$ with $0.2 < \delta < 0.5$). To compute the supremum functional, we picked the maximum of $|f_n(t) - f(t)|$ over a grid of 50 equally spaced values of t in the interval (0,1). Furthermore, for each sample size n and each value of h_n, we also computed $B = 1000$ copies of the weighted bootstrap statistics

$$Y_{nn} := \sqrt{2\delta\,\log n}\,\left(\frac{M_{nn}}{\lambda^{1/2}} - d_n\right) \quad \text{and} \quad \widehat{Y}_{nn} := \sqrt{2\delta\,\log n}\,\left(\frac{\widehat{M}_{nn}}{\lambda^{1/2}} - d_n\right), \tag{2.6}$$

where the weights $\epsilon_1, \cdots, \epsilon_n$ used in (2.1) were standard normal random variables. Also, for each value of h_n (and n), we computed $B = 1000$ copies of

$$Y_n^* := \sqrt{2\delta\,\log n}\,\left(\frac{M_n^*}{\lambda^{1/2}} - d_n\right) \quad \text{and} \quad \widehat{Y}_n^* := \sqrt{2\delta\,\log n}\,\left(\frac{\widehat{M}_n^*}{\lambda^{1/2}} - d_n\right);$$

these are Efron's original bootstrap counterparts of Y_n and $\widehat{Y}_n$. This entire process was then repeated 1000 times. To summarize our findings, let

$$U = \exp\Big(-2\exp(-Y_n)\Big) \quad \text{and} \quad \widehat{U} = \exp\Big(-2\exp(-\widehat{Y}_n)\Big)$$

and note that if n is not 'too small' then, by Theorem 1, each of the random variables U and $\widehat{U}$ should have, approximately, a Unif (0,1) distribution. Similarly, by Theorem 2, the random variables

$$V = B^{-1}\sum_{j=1}^{B} I\left\{Y_{nn,j} \le Y_n\right\} \quad \text{and} \quad \widehat{V} = B^{-1}\sum_{j=1}^{B} I\left\{\widehat{Y}_{nn,j} \le Y_n\right\}$$

should each be, approximately, a Unif (0,1) random variable; here, for $j = 1, \cdots, B = 1000$, the random variable $Y_{nn,j}$ (equivalently $\widehat{Y}_{nn,j}$) is the jth copy of Y_{nn} (equivalently $\widehat{Y}_{nn}$) in (2.6), based on the the jth sample of weights $\epsilon_{1,j}, \cdots, \epsilon_{n,j}$. Similarly, the random variables

$$W = B^{-1} \sum_{j=1}^{B} I\left\{Y_{n,j}^{*} \le Y_n\right\} \quad \text{and} \quad \widehat{W} = B^{-1} \sum_{j=1}^{B} I\left\{\widehat{Y}_{n,j}^{*} \le Y_n\right\}$$

should each be approximately a Unif (0,1) random variable, where $Y_{n,j}^{*}$ (equivalently $\widehat{Y}_{n,j}^{*}$) is the jth copy of Y_n^{*} (equivalently $\widehat{Y}_n^{*}$), computed based on the jth bootstrap sample $X_{1,j}^{*}, \cdots, X_{n,j}^{*}$; see (1.3) and the remarks after. Carrying out 1000 such Monte Carlo runs resulted in $U_1, \cdots, U_{1000}$; $\widehat{U}_1, \cdots, \widehat{U}_{1000}$; $V_1, \cdots, V_{1000}$; $\widehat{V}_1, \cdots, \widehat{V}_{1000}$; $W_1, \cdots, W_{1000}$, and $\widehat{W}_1, \cdots, \widehat{W}_{1000}$. Figure 1 gives the plots of the empirical distribution functions of V_i's, W_i's, and U_i's when the sample size is $n = 50$. The 45° line represents the true cdf of the Unif(0,1) distribution.

Plots (a), (b), and (c) show that the weighted bootstrap approximation performs much better than the large-sample theory (in the sense of capturing the true distribution of Y_n). This is reflected by the fact that in (a), (b), and (c), the empirical cdf of V_i's nearly coincides with the 45° line, for various choices of h_n. Similarly, the same conclusion applies to the usual (Efron's) bootstrap approximation, as shown by plots (d), (e), and (f) of Figure 1. Plots (g), (h), and (i) of Figure 1 show that U_i's are perhaps far from being Unif(0,1); this is in line with the well-known fact that the rate of convergence in Theorem 1 is very slow (logarithmic rate). Figure 2 gives the same plots for the empirical cdf of $\widehat{V}_i$'s (in plots (a), (b), (c)), and the empirical cdf of $\widehat{W}_i$'s (in plots (d), (e), (f)), and those of $\widehat{U}_i$'s (plots (g), (h), (i)). Once again, these plots confirm that the weighted bootstrap approximation (as well as Efron's original bootstrap) outperforms the large-sample theory (in the sense of capturing the true distribution of $\widehat{Y}_n$). Similar results are obtained for the case of $n = 100$ in Figure 3 and Figure 4.

Next, as a more formal approach, we carried out tests of hypothesis for the distributions of the resulting 1000 random variables ($U_1, \cdots, U_{1000}$; $\widehat{U}_1, \cdots, \widehat{U}_{1000}$; $\cdots$). 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 (=.10, .15, .20, .25, .30, .35, .40) the following tests of the null hypothesis were carried out.

1. The Kolmogorov-Smirnov tests:

   $H_0^{(1)}$: $U_1, \cdots, U_{300}$ are iid Unif(0,1)

   $H_0^{(2)}$: $V_1, \cdots, V_{300}$ are iid Unif(0,1)

   $H_0^{(3)}$: $W_1, \cdots, W_{300}$ are iid Unif(0,1)

   $H_0^{(4)}$: $\widehat{U}_1, \cdots, \widehat{U}_{300}$ are iid Unif(0,1)

   $H_0^{(5)}$: $\widehat{V}_1, \cdots, \widehat{V}_{300}$ are iid Unif(0,1)

$H_0^{(6)}$: $\widehat{W}_1, \cdots, \widehat{W}_{300}$ are iid Unif(0,1).

2. The Shapiro-Wilk tests (of normality):
   $H_0^{(7)}$: $\Phi^{-1}(U_1), \cdots, \Phi^{-1}(U_{300})$ are iid N(0,1)
   $H_0^{(8)}$: $\Phi^{-1}(V_1), \cdots, \Phi^{-1}(V_{300})$ are iid N(0,1)
   $H_0^{(9)}$: $\Phi^{-1}(W_1), \cdots, \Phi^{-1}(W_{300})$ are iid N(0,1)
   $H_0^{(10)}$: $\Phi^{-1}(\widehat{U}_1), \cdots, \Phi^{-1}(\widehat{U}_{300})$ are iid N(0,1)
   $H_0^{(11)}$: $\Phi^{-1}(\widehat{V}_1), \cdots, \Phi^{-1}(\widehat{V}_{300})$ are iid N(0,1)
   $H_0^{(12)}$: $\Phi^{-1}(\widehat{W}_1), \cdots, \Phi^{-1}(\widehat{W}_{300})$ are iid N(0,1),

where Φ is the cdf of the standard normal distribution. The p-values corresponding to the hypotheses $H_0^{(k)}$, $k = 2, 3, 5, 6, 8, 9, 11, 12$ were all larger than 5% (and in fact, in most cases larger than 10%), indicating that both weighted and regular bootstraps work at 5% significance level. This was true for both n= 50 and 100. On the other hand, all the other p-values (i.e., the p-values for $H_0^{(k)}$, $k = 1, 4, 7, 10$) were virtually less than 10^{-4}.

References

[1] Barbe, P. and P. Bertail. (1995). *The weighted bootstrap.* Lecture Notes in Statistics, 98. Springer-Verlag, New York.

[2] Bickel, P. and M. Rosenblatt. (1973). On some global measures of the deviations of density function estimates. *Annals of Statistics,* **1**, 1071-1095.

[3] Burke, M. (2000). Multivariate tests-of-fit and uniform confidence bands using a weighted bootstrap. *Statistics and Probability Letters,* **46**, 13-20.

[4] Csörgő, M., L. Horvath, and P. Kokoszka. (2000). Approximation for bootstrapped empirical processes. *Proceedings of the American Mathematical Society.* **128**, 2457-2464.

[5] Devroye, L. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canadian Journal of Statistics,* **6**, 179-191.

[6] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics,* **7**, 1-26.

[7] Hall, P. and E. Mammen. (1994). On general resampling algorithms and their performance in distribution estimation. *Annals of Statistics,* **22**, 2011-2030.

[8] Hall, P. (1991). On convergence rates of suprema. *Probability Theory and Related Fields,* **89**, 447-455.

[9] Horvath, L. (1991). On L_p-norms of multivariate density estimators. *Annals of Statistics,* **19**, 1933-1949.

[10] Horvath, L., P. Kokoszka, and J. Steinebach. (2000). Approximations for weighted bootstrap processes with an application. *Statistics and Probability Letters,* **48**, 59-70.

[11] Konakov, V. and V. Piterbarg. (1984). On the convergence rate of maximal deviation distribution for kernel regression estimates. *Journal of Multivariate Analysis,* **15**, 279-294.

[12] Mason, D. and M. Newton. (1992). A rank statistics approach to the consistency of a general bootstrap. *Annals of Statistics,* **20**, 1611-1624.

[13] Mojirsheibani, M. (2007). Some approximations to L_p-statistics of kernel density estimators. *Statistics,* **41**, 203-220.

[14] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics,* **33**, 1065-1076.

[15] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics,* **27** 832-837.

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

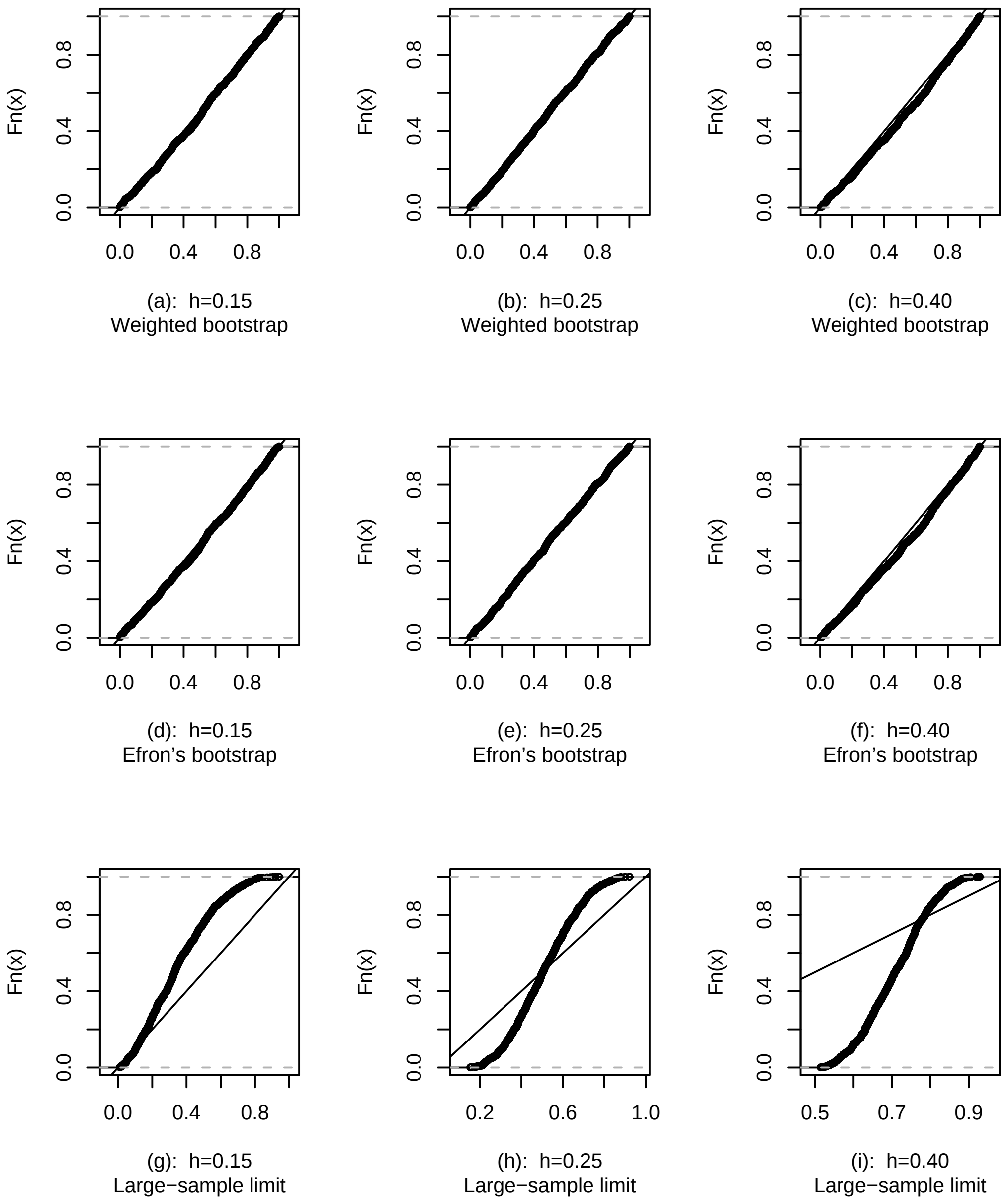

Figure 1: Plots of the empirical cdf's (n=50) of V_i's appear in (a), (b), and (c), of W_i's are in (d), (e), and (f), and of U_i's appear in (g), (h), and (i), for different values of h_n.

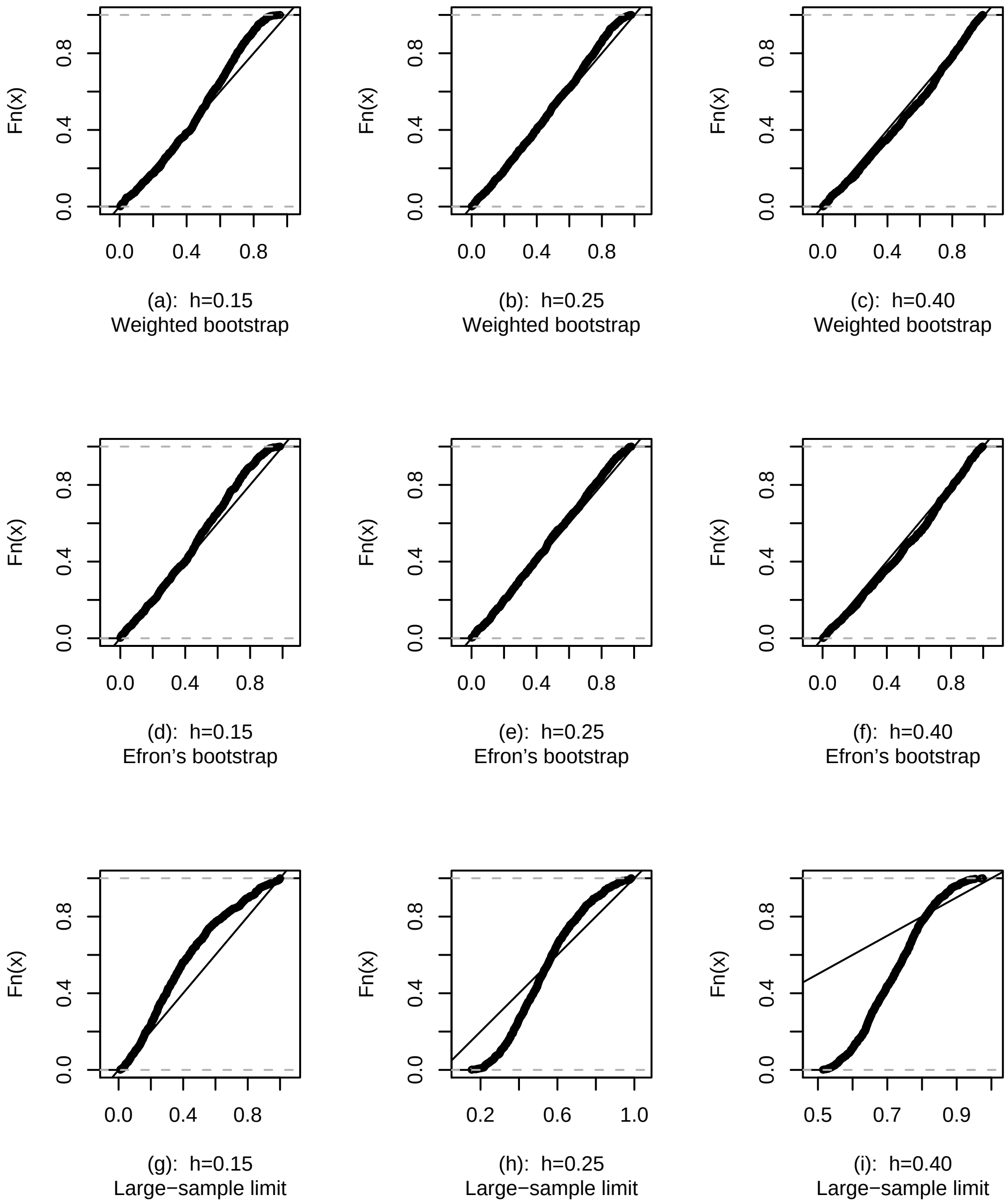

Figure 2: Plots of the empirical cdf's (n=50) of $\widehat{V}_i$'s appear in (a), (b), and (c), of $\widehat{W}_i$'s are in (d), (e), and (f), and of $\widehat{U}_i$'s appear in (g), (h), and (i), for different values of h_n.

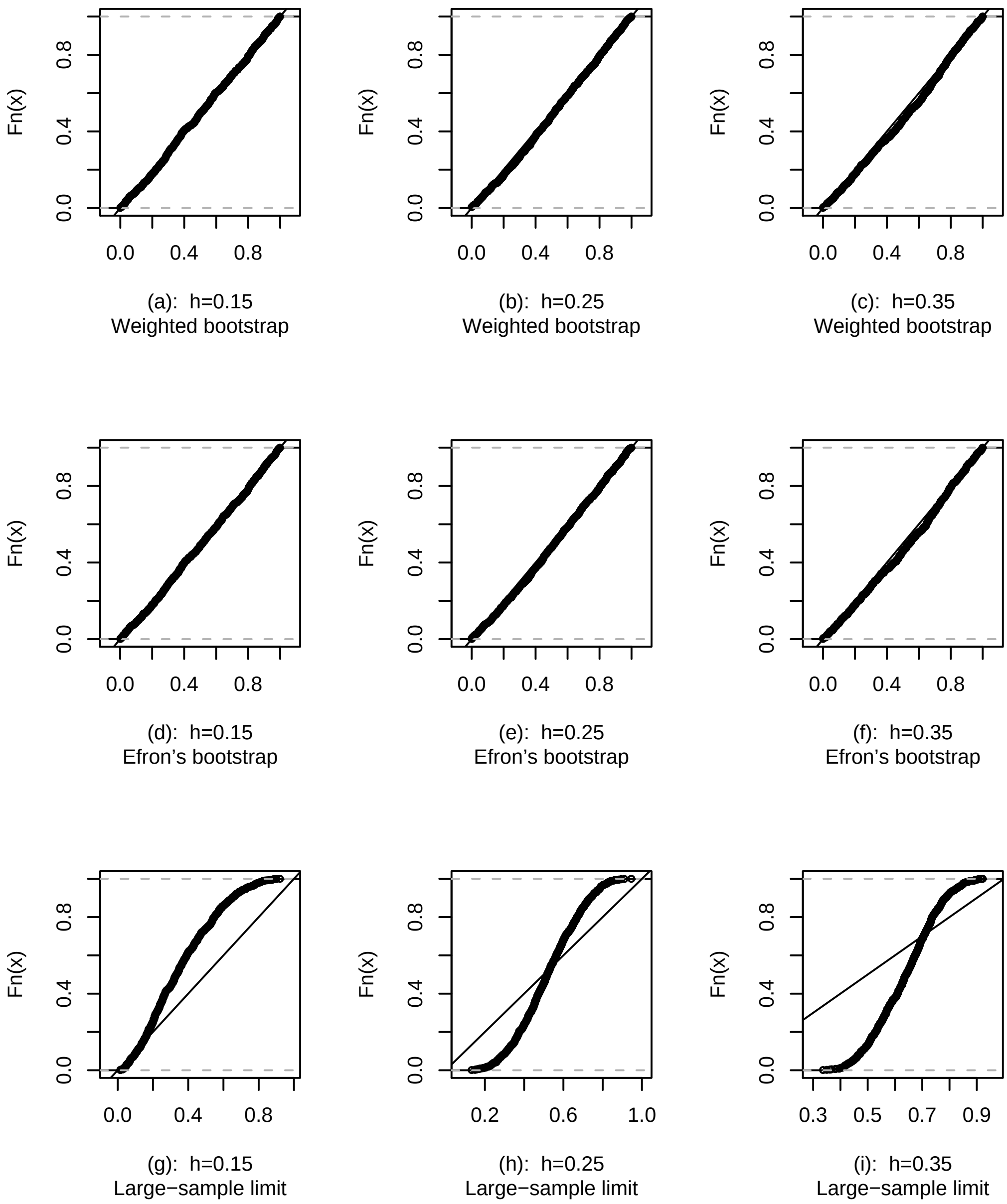

Figure 3: Plots of the empirical cdf's (n=100) of V_i's appear in (a), (b), and (c), of W_i's are in (d), (e), and (f), and of U_i's appear in (g), (h), and (i), for different values of h_n.

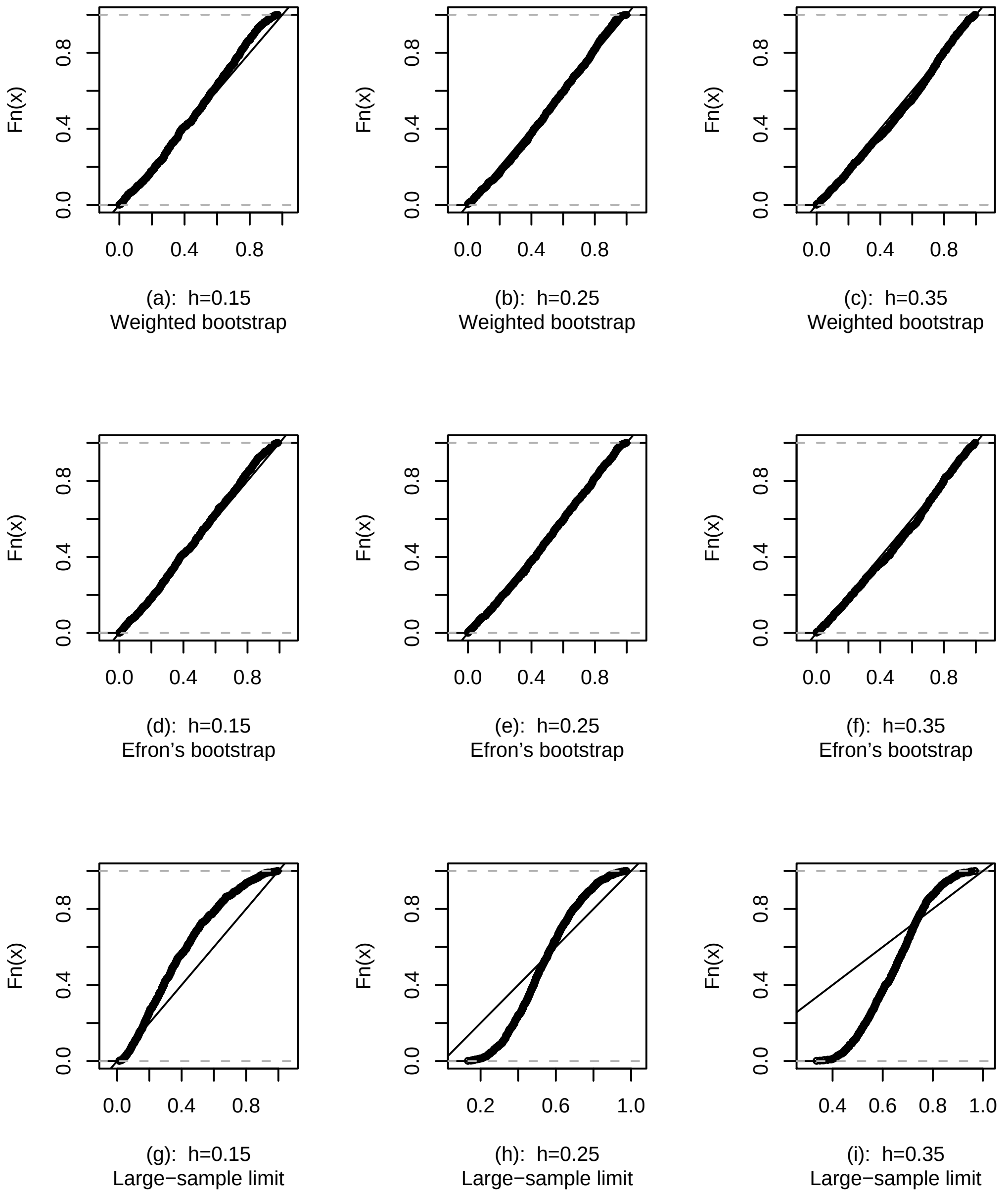

Figure 4: Plots of the empirical cdf's (n=100) of $\widehat{V}_i$'s appear in (a), (b), and (c), of $\widehat{W}_i$'s are in (d), (e), and (f), and of $\widehat{U}_i$'s appear in (g), (h), and (i), for different values of h_n.