A note on the performance of bootstrap kernel density estimation with small re-sample sizes

Majid Mojirsheibani[1]
Department of Mathematics, California State University Northridge, CA, 91330, USA

Abstract

This paper studies the unconditional limiting distribution of the maximal deviation of bootstrap kernel density estimators with re-sample sizes that are different from the sample size, n. More specifically, we study the convergence rates of such statistics when the bootstrap sample size may be orders of magnitude smaller than n. An application to big-data scenarios is given.

Keywords: Kernel, bootstrap, Brownian bridge, approximation.

1 Introduction

Let $X_1, \cdots, X_n$ be n independent and identically distributed (i.i.d.) random variables with the underlying distribution function F and the probability density function $f = F'$. Also, let $f_n(x) = (nh)^{-1} \sum_{i=1}^n K((x - X_i)/h)$ be the popular Parzen-Rosenblatt kernel estimator of f (Parzen (1962), Rosenblatt (1956)), where K is the kernel used with the smoothing parameter $h \equiv h(n)$. The asymptotic distribution of the maximal deviation of $f_n(x)$ from $f(x)$, which plays a crucial role in statistical inference (as in goodness-of-fit tests for f or the construction of uniform confidence bands for f over compact sets), follows from the classical results of Bickel and Rosenblatt (1973), Konakov and Piterbarg (1984), Rio (1994), and Muminov (2011, 2012), among others, who studied the limiting distribution of the properly normalized versions of the statistic $\sup_{0\leq t\leq 1} |f_n(t) - f(t)|/\sqrt{f(t)}$; here, the interval $[0, 1]$ can be replaced by any compact set. Given the slow rate of convergence (logarithmic only) of this statistic to its limiting distribution (Konakov and Piterbarg (1984), one can always consider the bootstrap methodology as an alternative approximation; see, for example, Mojirsheibani (2012) and Al-Sharadqah et al. (2020).

The focus of this paper is to take a closer look at the asymptotic superiority of such bootstrap approximations when the bootstrap re-sample size is substantially smaller than the original sample size, n. This approach, which may be viewed as a *virtual* bootstrap, can be particularly beneficial in big-data scenarios where the data size n may be huge. Drawing bootstrap re-samples of smaller sizes may resemble the m out of n bootstrap (Bickel et al. (1997)), but the latter method is usually intended to remedy the situations where Efron's (1979) original algorithm fails (e.g., the distribution of the largest order statistic); the following quote from Bickel and Sakov (2008, p. 967) asserts this:

> "The choice of m can be crucial, and two issues are involved. The first is that the user does not know, a-priori, whether the bootstrap works or not, in his case. The second is the choice of m, in case of n-bootstrap failure."

In contrast, here we already know that the bootstrap works for the statistic of interest, but the sample size is far too large to draw repeated bootstrap samples of size n. In the next section we show that even if the bootstrap sample size m is orders of magnitude smaller than n, the bootstrap

[1]Corresponding author. Email: majid.mojirsheibani@csun.edu

approximation can still have a fast rate of convergence (instead of logarithmic). Our results can also provide some partial guidance in choosing the bootstrap sample size m. The proofs of our main results employ tools from strong approximation theory.

2 Main results

2.1 The setup and the background

Once again, let $f_n(x) = (nh)^{-1}\sum_{i=1}^n K((x-X_i)/h)$ be the kernel density estimator of f, where X_i, $i = 1,\ldots,n$ are i.i.d with the unknown density f. We also state a number of assumptions.

Assumption (A). The function $f^{1/2}(t)$, $a \le t \le b$, is strictly positive and satisfies the Lipschitz condition of order 1.

Assumption (B). There exists an integer $s \ge 1$ and an $\varepsilon > 0$ such that the density f has partial derivatives of up to order s that are bounded on $(a-\varepsilon,\, b+\varepsilon)$.

Assumption (C). The kernel K is finite and satisfies $\int K(u)du = 1$, $K'(u)$ is continuous and $\int (K'(u))^2 du < \infty$. Moreover, $\int u^r K(u)du = 0$ for all $r \in \{1,\ldots,s\}$, i.e., K is a kernel of order $s+1$, where s is as in *Assumption (B).*

Now, for $-\infty < a \le t \le b < \infty$, consider the statistics

$$\Gamma_n(t) = \sqrt{nh}\,\big[f_n(t) - E\big(f_n(t)\big)\big]/\sqrt{f(t)}, \quad \text{and} \quad \Gamma_n^*(t) = \sqrt{nh}\,\big[f_n(t) - f(t)\big]/\sqrt{f(t)}. \tag{1}$$

The statistic $\Gamma_n(t)$ in (1) plays a central role in the papers cited in Section 1. In fact, the limiting distribution of $\Gamma_n^*(t)$ follows from that of $\Gamma_n(t)$ because the stated assumptions ensure that the bias term $E\big(f_n(t)\big) - f(t)$ goes to zero fast enough. The limiting distribution of the properly centered and normalized versions of these statistics have been studied by Bickel and Rosenblatt (1973), Konakov and Piterbarg (1984), Rio (1994), and Muminov (2011, 2012), and others, and is shown to be a double-exponential distribution. In particular, the following is due to Konako and Piterbarg (1984):

Theorem 1 *Let $\Gamma_n(t)$ and $\Gamma_n^*(t)$ be as in (1) and suppose that assumptions (A), (B), and (C) hold. Let $h = n^{-\delta}$, for any $(1+2s)^{-1} < \delta < 0.5$, where s is as in assumption (C). Then, one has*

$$\lim_{n\to\infty} P\Big\{\ell_h\, \lambda^{-1/2} \sup_{a\le t\le b} \big|\Gamma_n^*(t)\big| - \ell_h^2 \;\le\; x\Big\} \;=\; \exp\{-2\exp(-x)\}, \tag{2}$$

$$\lim_{n\to\infty} P\Big\{\ell_h\, \lambda^{-1/2} \sup_{a\le t\le b} \big|\Gamma_n(t)\big| - \ell_h^2 \;\le\; x\Big\} \;=\; \exp\{-2\exp(-x)\}, \tag{3}$$

where

$$\lambda = \int K^2(u)\,du \quad \text{and} \quad \ell_h \;=\; \sqrt{2\log\Big(\frac{b-a}{h}\Big) + \log\Big(\frac{1}{\lambda}\int (K'(u))^2\,du\Big) + 2\log\Big(\frac{1}{2\pi}\Big)}. \tag{4}$$

To present our results, let f_{mn} be the bootstrap version of the kernel density estimator f_n, i.e.,

$$f_{mn}(t) = (mh)^{-1}\sum_{i=1}^m K((t-X_i^*)/h), \tag{5}$$

where $(X_1^*, \ldots, X_m^*)$ is a sample of size m drawn with replacement from the original sample; thus, $X_1^*, \ldots, X_m^*$ are conditionally independent (conditional on $X_1, \cdots, X_n$). Now, consider the following statistic which is the bootstrap counterpart of $\Gamma_n(t)$ in (1)

$$\Gamma_{mn}(t) = \sqrt{mh}\,\big[f_{mn}(t) - f_n(t)\big]\big/\sqrt{f_n(t)}\,, \quad a \le t \le b. \tag{6}$$

Also, consider the following "studentized" counterpart of (1)

$$\widehat{\Gamma}_n(t) = \sqrt{nh}\,\big[f_n(t) - E\big(f_n(t)\big)\big]\big/\sqrt{f_n(t)}\,, \quad a \le t \le b \tag{7}$$

and its bootstrap version

$$\widehat{\Gamma}_{mn}(t) = \sqrt{mh}\,\big[f_{mn}(t) - f_n(t)\big]\big/\sqrt{f_{mn}(t)}\,, \quad a \le t \le b. \tag{8}$$

Define the random variables

$$M_n = \ell_h\,\lambda^{-1/2} \sup_{a\le t\le b} \big|\Gamma_n(t)\big| - \ell_h^2 \qquad \widehat{M}_n = \ell_h\,\lambda^{-1/2} \sup_{a\le t\le b} \big|\widehat{\Gamma}_n(t)\big| - \ell_h^2 \tag{9}$$

$$M_{mn} = \ell_h\,\lambda^{-1/2} \sup_{a\le t\le b} \big|\Gamma_{mn}(t)\big| - \ell_h^2 \qquad \widehat{M}_{mn} = \ell_h\,\lambda^{-1/2} \sup_{a\le t\le b} \big|\widehat{\Gamma}_{mn}(t)\big| - \ell_h^2 \tag{10}$$

where λ and ℓ_h are as in (4). It is not difficult to show that the bootstrap statistics M_{mn} and $\widehat{M}_{mn}$ work in the sense that their unconditional limiting distributions are the same as those in Theorem 1. In fact, under the conditions of Theorem 2, one has $M_{mn} \to^d Y$ and $\widehat{M}_{mn} \to^d Y$, where $P\{Y \le y\} = \exp\{-2\,\exp(-y)\}$, $\forall y \in \mathbb{R}$. The proof of such results rely on the strong approximation of empirical and bootstrapped empirical processes by sequences of Brownian bridges. Here, we prove a stronger result showing that even if m is orders of magnitude smaller than n (such as $m = \sqrt{n}$), the bootstrap approximation will still enjoy a polynomial rate of convergence (instead of logarithmic). This can be particularly useful in big-data scenarios since it can alleviate the formidable computational cost of drawing bootstrap samples of size n, while still retaining the benefits of bootstrap methodology.

Theorem 2 *Let M_n, $\widehat{M}_n$, M_{mn}, and $\widehat{M}_{mn}$ be as (9) and (10), and suppose that assumptions (A), (B), and (C) hold. Let $h = n^{-\delta}$ and $m = n^{\nu}$ for any $(1+2s)^{-1} < \delta < 1/3$ and any $\delta < \nu < 1$, where s is as in assumption (C). Then, one has*

(i) $$\sup_{-\infty<x<\infty} \Big| P\Big\{\widehat{M}_{mn} \le x\Big\} - P\Big\{\widehat{M}_n \le x\Big\}\Big| = \mathcal{O}\big(n^{-\lambda}\,(\log n)^{3/2}\big) + \mathcal{O}(n^{-\beta})$$

and

(ii) $$\sup_{-\infty<x<\infty} \Big| P\big\{M_{mn} \le x\big\} - P\big\{M_n \le x\big\}\Big| = \mathcal{O}\big(n^{-\lambda}\,(\log n)^{3/2}\big) + \mathcal{O}(n^{-\beta}),$$

where λ and β are positive constants not depending on m or n.

In what follows, we will assume, without loss of generality, that all random variables and precesses are defined on the same probability space; for more on this one may refer to Section A.2 of Csörgő and Horváth (1993). To proceed, first we state some preliminary results. For $t \in \mathbb{R}$, let

$$F_n(t) = n^{-1}\sum_{i=1}^{n} I\{X_i \le t\}, \qquad F_{mn}(t) = m^{-1}\sum_{i=1}^{m} I\{X_i^* \le t\} \tag{11}$$

$$\beta_n(t) = n^{1/2}(F_n(t) - F(t)), \qquad \beta_{mn}(t) = m^{1/2}(F_{mn}(t) - F_n(t)) \tag{12}$$

First we state a result on the best approximation of bootstrapped empirical processes by a sequence of Brownian bridges, due to Csörgő et al. (1999), (for another closely related important result along these lines, one may refer to Alvarez-Andrade and Bouzebda (2013)).

Lemma 1 *Let $\beta_{mn}(t)$ be the bootstrap empirical process defined in (12). Then one can define a sequence of Brownian bridges $\{\mathbb{B}_{mn}(t)\,,\ 0 \le t \le 1\}$ such that*

$$P\Big\{ \sup_{-\infty<t<\infty} \Big|\beta_{mn}(t) - \mathbb{B}_{mn}(F(t))\Big| \;>\; \sqrt{m}\big(c_1 \log m + x\big) + \sqrt{n}\big(c_4 \log n + y\big) \Big\} \;\le\; c_2 e^{-c_3 x} + c_5 e^{-c_6 y},$$

for all $x, y > 0$, where $c_1, \ldots, c_6$ are positive constants.

Next, we state an inequality which will be useful in the sequel.

Lemma 2 *Let X and Y be any random variables. Then for all $\varepsilon > 0$ and every real u*

$$\Big|P\left\{|X| \le u\right\} - P\left\{|Y| \le u\right\}\Big| \quad\le\quad P\left\{|X - Y| \ge \varepsilon\right\} + P\Big\{\Big||Y| - u\Big| < \varepsilon\Big\}.$$

PROOF OF THEOREM 2
Proof of part (i). Start by defining

$$\widetilde{M}_{mn} \;=\; \ell_h\, \lambda^{-1/2} h^{-1/2} \sup_{a\le t\le b} \left| \frac{1}{\sqrt{f(t)}} \int K((t-s)/h)\, d\,\mathbb{B}_{mn}(F(s)) \right| - \ell_h^2\,, \tag{13}$$

where $\mathbb{B}_{mn}(\cdot)$ is as in Lemma 1, and observe that for each real x and any constant $\varepsilon_{mn} > 0$, (where ε_{mn} can depend on m and n),

$$\begin{aligned}
&\big|P\{\widehat{M}_{mn} \le x\} - P\{\widetilde{M}_{mn} \le x\}\big| \\
&= \Big|P\Big\{\ell_h \lambda^{-1/2} \sup_{a\le t\le b} \Big|\widehat{\Gamma}_{mn}(t)\Big| \le x + \ell_h^2\Big\} \\
&\quad - P\Big\{\ell_h \lambda^{-1/2} h^{-1/2} \sup_{a\le t\le b} \Big|\frac{1}{\sqrt{f(t)}} \int K((t-s)/h)\, d\,\mathbb{B}_{mn}(F(s))\Big| \le x + \ell_h^2\Big\}\Big| \\
&\le P\Big\{\ell_h \lambda^{-1/2} \Big| \sup_{a\le t\le b} \Big|\widehat{\Gamma}_{mn}(t)\Big| - h^{-1/2} \sup_{a\le t\le b} \Big|\frac{1}{\sqrt{f(t)}} \int K((t-s)/h)\, d\,\mathbb{B}_{mn}(F(s))\Big|\Big| \ge \varepsilon_{mn}\Big\} \\
&\quad + P\Big\{\Big|\ell_h \lambda^{-1/2} h^{-1/2} \sup_{a\le t\le b} \Big|\frac{1}{\sqrt{f(t)}} \int K((t-s)/h)\, d\,\mathbb{B}_{mn}(F(s))\Big| - x - \ell_h^2\Big| < \varepsilon_{mn}\Big\}, \\
&\qquad \big(\text{which follows from Lemma 2, for any constants } \varepsilon_{mn} > 0\big) \\
&=: S_{n,1} + S_{n,2}(x).
\end{aligned} \tag{14}$$

Now observe that for any constants $\varepsilon'_{mn} > 0$ and $\varepsilon''_{mn} > 0$ satisfying $\varepsilon'_{mn} + \varepsilon''_{mn} = \varepsilon_{mn}$, one has

$$\begin{aligned}
S_{n,1} \;&\le\; P\Big\{\ell_h \lambda^{-1/2} \sup_{a\le t\le b} \Big|\widehat{\Gamma}_{mn}(t) - \frac{h^{-1/2}}{\sqrt{f(t)}} \int K((t-s)/h)\, d\,\mathbb{B}_{mn}(F(s))\Big| \ge \varepsilon_{mn}\Big\} \\
&=\; P\Big\{ \sup_{a\le t\le b} \Big| \frac{1}{\sqrt{f_{mn}(t)}} \int K((t-s)/h)\, d\beta_{mn}(s) - \frac{1}{\sqrt{f(t)}} \int K((t-s)/h)\, d\,\mathbb{B}_{mn}(F(s))\Big| \\
&\qquad\qquad \ge \frac{\lambda^{1/2} h^{1/2} \varepsilon_{mn}}{\ell_h}\Big\} \\
&\le\; P\Big\{ \sup_{a\le t\le b} \frac{1}{\sqrt{f_{mn}(t)}} \Big| \int K((t-s)/h)\, d\beta_{mn}(s) - \int K((t-s)/h)\, d\,\mathbb{B}_{mn}(F(s))\Big| \ge \frac{\lambda^{1/2} h^{1/2} \varepsilon'_{mn}}{\ell_h}\Big\}
\end{aligned}$$

$$+ P\Big\{ \sup_{a\le t\le b} \Big| \Big[\frac{1}{\sqrt{f_{mn}(t)}} - \frac{1}{\sqrt{f(t)}} \Big] \int K((t-s)/h)\, d\,\mathbb{B}_{mn}(F(s)) \Big| \ge \frac{\lambda^{1/2} h^{1/2} \varepsilon''_{mn}}{\ell_h} \Big\}$$

$\big($where the choices of ε'_{mn} and ε''_{mn} will be given later$\big)$

$$:= \quad S_{n,1}(i) + S_{n,1}(ii). \tag{15}$$

However, the term $S_{n,1}(i)$ in (15) can be bounded as follows. Let $\mu_K = \big|\int dK(u)\big|$, then

$$\begin{aligned} S_{n,1}(i) &= P\Big\{ \sup_{a\le t\le b} \Big| \frac{1}{\sqrt{f_{mn}(t)}} \int \Big[\mathbb{B}_{mn}\big(F(t-uh)\big) - \beta_{mn}(t-uh) \Big]\, dK(u) \Big| \ge \frac{\lambda^{1/2} h^{1/2} \varepsilon'_{mn}}{\ell_h} \Big\} \\ &\le P\Big\{ \frac{\mu_K}{\inf_{a\le t\le b} \sqrt{f_{mn}(t)}} \cdot \sup_{-\infty<v<\infty} \Big| \beta_{mn}(v) - \mathbb{B}_{mn}(F(v)) \Big| \ge \frac{\lambda^{1/2} h^{1/2} \varepsilon'_{mn}}{\ell_h} \Big\} \\ &\le P\Big\{ \Big[\frac{\mu_K}{\sqrt{f_0/2}} \sup_{-\infty<v<\infty} \Big| \beta_{mn}(v) - \mathbb{B}_{mn}(F(v)) \Big| \ge \frac{\lambda^{1/2} h^{1/2} \varepsilon'_{mn}}{\ell_h} \Big] \\ &\qquad \cap \Big[\inf_{a\le t\le b} f_{mn}(t) \ \ge\ f_0/2 \Big] \Big\} + P\Big\{ \inf_{a\le t\le b} f_{mn}(t) \ <\ f_0/2 \Big\}. \end{aligned} \tag{16}$$

It is straightforward to show that

$$\begin{aligned} &P\Big\{ \inf_{a\le t\le b} f_{mn}(t) < f_0/2 \Big\} \\ &\le P\Big\{ \sup_{a\le t\le b} \big| f_{mn}(t) - f_n(t) \big| > \frac{f_{\max}}{2} - \frac{f_0}{4} \Big\} + P\Big\{ \sup_{a\le t\le b} \big| f_n(t) - f(t) \big| > \frac{f_{\max}}{2} - \frac{f_0}{4} \Big\}, \end{aligned} \tag{17}$$

where $f_{\max} := \sup_{a\le t\le b} f(t) < \infty$ by Assumption (B). Furthermore, one can show that

$$\sup_{a\le t\le b} \big| f_{mn}(t) - f_n(t) \big| \ \le\ m^{-1/2} \mu_K \cdot \Big(\sup_{-\infty<v<\infty} \Big| \beta_{mn}(v) - \mathbb{B}_{mn}(F(v)) \Big| + \sup_{0\le s\le 1} |\mathbb{B}_{mn}(s)| \Big), \tag{18}$$

where $\mathbb{B}_{mn}(\cdot)$ is as in Lemma 1. Thus, in view of (16), (17), and (18), one finds

$$\begin{aligned} S_{n,1}(i) \ &\le\ P\Big\{ \frac{\mu_K}{\sqrt{f_0/2}} \sup_{-\infty<v<\infty} \Big| \beta_{mn}(v) - \mathbb{B}_{mn}(F(v)) \Big| \ \ge\ \frac{\lambda^{1/2} h^{1/2} \varepsilon'_{mn}}{\ell_h} \Big\} \\ &\quad + P\Big\{ \sup_{a\le t\le b} \big| f_n(t) - f(t) \big| \ >\ \frac{f_{\max}}{2} - \frac{f_0}{4} \Big\} \\ &\quad + P\Big\{ \sup_{-\infty<v<\infty} \Big| \beta_{mn}(v) - \mathbb{B}_{mn}(F(v)) \Big| \ \ge\ \big(f_{\max} - \frac{f_0}{2}\big)\sqrt{m}/(4\mu_K) \Big\} \\ &\quad + P\Big\{ \sup_{0\le s\le 1} \Big| \mathbb{B}_{mn}(s) \Big| \ge (f_{\max} - \frac{f_0}{2})\sqrt{m}/(4\mu_K) \Big\} \\ &=: \ \Delta_n(1) + \Delta_n(2) + \Delta_n(3) + \Delta_n(4) \end{aligned} \tag{19}$$

Now, let $c_1, \ldots, c_6$ be the positive constants of Lemma 1 and put

$$\varepsilon'_{mn} = \frac{\ell_h\, \mu_K}{(f_0/2)^{1/2} \lambda^{1/2} h^{1/2}} \cdot \Big[\frac{(c_1 + b_1)\log m}{\sqrt{m}} + \frac{(c_4 + b_2)\log n}{\sqrt{n}} \Big], \tag{20}$$

where b_1 and b_2 are any constants satisfying $b_1 > \frac{1}{c_3}$ and $b_2 > \frac{1}{c_6}$. Then, by Lemma 1, we have

$$\Delta_n(1) \ \le\ c_2\, e^{-c_3 b_1 \log m} + c_5\, e^{-c_6 b_2 \log n}$$

$$= \; c_2\, m^{-c_3 b_1} + c_5\, n^{-c_6 b_2} \; = \; o(m^{-1}) + o(n^{-1}), \quad \big(\text{because } b_1 c_3 > 1 \text{ and } b_2 c_6 > 1\big) \tag{21}$$

As for the term $\Delta_n(3)$, since for m large enough $\frac{(f_{\max} - \frac{f_0}{2})\sqrt{m}}{4\mu_K} > \frac{(c_1+b_1)\log m}{\sqrt{m}} + \frac{(c_4+b_2)\log n}{\sqrt{n}}$, where b_1 and b_2 are as in (20), it immediately follows from Lemma 1 that for m large

$$\begin{aligned} \Delta_n(3) \;&\leq\; P\left\{ \sup_{-\infty<v<\infty} \left|\beta_{mn}(v) - \mathbb{B}_{mn}(F(v))\right| \geq \frac{(c_1+b_1)\log m}{\sqrt{m}} + \frac{(c_4+b_2)\log n}{\sqrt{n}} \right\} \\ &\leq\; c_2\, e^{-c_3 b_1 \log m} + c_5\, e^{-c_6 b_2 \log n} \;=\; o(m^{-1}) + o(n^{-1}). \end{aligned} \tag{22}$$

To bound the term $\Delta_n(4)$, let $\big\{\mathbb{B}(t),\ 0 \leq t \leq 1\big\}$ be a Brownian bridge and observe that

$$\big\{\mathbb{B}_{mn}(t),\ 0 \leq t \leq 1\big\} \overset{\mathcal{D}}{=} \big\{\mathbb{B}(t),\ 0 \leq t \leq 1\big\}, \quad \text{for each } m = 1, 2, \ldots, \text{ and } n = 1, 2, \ldots \tag{23}$$

Therefore, in view of the distribution of the maximal modulus of a Brownian bridge, one has

$$\begin{aligned} \Delta_n(4) \;&=\; P\Big\{ \sup_{0\leq s\leq 1} \left|\mathbb{B}(s)\right| \geq c_7\sqrt{m} \Big\}, \qquad \text{where } c_7 = \big(f_{\max} - \tfrac{f_0}{2}\big)\big/(4\mu_K) > 0 \\ &=\; 2\sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2 (c_7\sqrt{m})^2} \;=\; \mathcal{O}\left(e^{-c_8 m}\right), \qquad \text{where } c_8 = 2c_7^2. \end{aligned} \tag{24}$$

Putting together (21), (22), (24), and the fact that $\Delta_n(2) = \mathcal{O}(\exp(-c_9 n h^2))$, $\exists\, c_9 > 0$, one finds

$$S_{n,1}(i) = o(m^{-1}) + o(n^{-1}). \tag{25}$$

To deal with the term $S_{n,1}(ii)$ in (15), we start by writing

$$\begin{aligned} S_{n,1}(ii) \;&\leq\; P\Bigg\{ \sup_{a\leq t\leq b} \left| \frac{f_{mn}(t) - f(t)}{\sqrt{f_{mn}(t)}\, f(t) + f_{mn}(t)\sqrt{f(t)}} \right| \cdot \sup_{0\leq s\leq 1} \left|\mathbb{B}_{mn}(s)\right| \geq \frac{\lambda^{1/2} h^{1/2} \varepsilon''_{mn}}{\ell_h\, \mu_K} \\ &\qquad\qquad \cap \left[\inf_{a\leq t\leq b} f_{mn}(t) \geq \frac{f_0}{2} \right] \Bigg\} + P\left\{ \inf_{a\leq t\leq b} f_{mn}(t) < \frac{f_0}{2} \right\}. \\ &=:\; S'_{n,1}(ii) + S''_{n,1}(ii). \end{aligned} \tag{26}$$

But, in view of (17) and (18), and with $\Delta_n(2)$, $\Delta_n(3)$, and $\Delta_n(4)$ as in (19), one immediately finds

$$S''_{n,1}(ii) \;\leq\; \Delta_n(2) + \Delta_n(3) + \Delta_n(4) \;=\; o(m^{-1}) + o(n^{-1}). \tag{27}$$

Furthermore, for any $\varepsilon''_{mn,1} > 0$ and $\varepsilon''_{mn,2} > 0$ satisfying $\varepsilon''_{mn,1} + \varepsilon''_{mn,2} = \varepsilon''_{mn}$, one has

$$\begin{aligned} S'_{n,1}(ii) \;&\leq\; P\Bigg\{ \sup_{a\leq t\leq b} |f_{mn}(t) - f(t)| \cdot \sup_{0\leq s\leq 1} \left|\mathbb{B}_{mn}(s)\right| \geq \frac{C_{11}\,\lambda^{1/2} h^{1/2} \varepsilon''_{mn}}{\ell_h\, \mu_K} \Bigg\} \\ &\qquad\qquad \text{where } C_{11} > 0 \text{ can be taken to be } \sqrt{f_0/2}\, f_0 + (f_0/2)\sqrt{f_0} \\ &\leq\; P\Bigg\{ \sup_{a\leq t\leq b} |f_{mn}(t) - f_n(t)| \cdot \sup_{0\leq s\leq 1} \left|\mathbb{B}_{mn}(s)\right| \geq \frac{C_{11}\,\lambda^{1/2} h^{1/2} \varepsilon''_{mn,1}}{\ell_h\, \mu_K} \Bigg\} \\ &\qquad + P\Bigg\{ \sup_{a\leq t\leq b} |f_n(t) - f(t)| \cdot \sup_{0\leq s\leq 1} \left|\mathbb{B}_{mn}(s)\right| \geq \frac{C_{11}\,\lambda^{1/2} h^{1/2} \varepsilon''_{mn,2}}{\ell_h\, \mu_K} \Bigg\} \\ &=:\; T_{mn}(i) + T_{mn}(ii) \end{aligned} \tag{28}$$

Now, in view of (18), we find

$$
\begin{aligned}
T_{mn}(i) \quad &\leq \quad P\Big\{ \sup_{a\leq u\leq b} \Big|\beta_{mn}(u) - \mathbb{B}_{mn}(F(u))\Big| \cdot \sup_{0\leq s\leq 1} \big|\mathbb{B}_{mn}(s)\big| \geq \frac{C_{11}\,\lambda^{1/2}h^{1/2}m^{1/2}\,\varepsilon''_{mn,1}}{2\,\ell_h\,\mu_K} \Big\} \\
&\quad + P\Big\{ \Big(\sup_{0\leq s\leq 1} \Big|\mathbb{B}_{mn}(s)\Big| \Big)^2 \geq \frac{C_{11}\,\lambda^{1/2}h^{1/2}m^{1/2}\,\varepsilon''_{mn,1}}{2\,\ell_h\,\mu_K} \Big\} \\
&=: \quad T'_{mn}(i) + T''_{mn}(i). \qquad (29)
\end{aligned}
$$

Therefore, choosing

$$
\varepsilon''_{mn,1} = \frac{2\,\ell_h\,\mu_K\,\log m}{C_{11}\,\lambda^{1/2}h^{1/2}m^{1/2}} \qquad (30)
$$

one finds

$$
T'_{mn}(i) \quad \leq \quad P\Big\{ \sup_{a\leq u\leq b} \Big|\beta_{mn}(u) - \mathbb{B}_{mn}(F(u))\Big| \cdot \sqrt{\log m} \;\geq\; \log m \Big\} + P\Big\{ \sup_{0\leq s\leq 1} \Big|\mathbb{B}_{mn}(s)\Big| \geq \sqrt{\log m} \Big\}
$$

However, $\sqrt{\log m} > \frac{(c_1+b_1)\log m}{\sqrt{m}} + \frac{(c_4+b_2)\log n}{\sqrt{n}}$ for large m, where $b_1 > \frac{1}{c_3}$ and $b_2 > \frac{1}{c_6}$. Thus by Lemma 1, for large m, $P\Big\{ \sup_{a\leq u\leq b} \big|\beta_{mn}(u) - \mathbb{B}_{mn}(F(u)) \;\geq\; \sqrt{\log m} \Big\} \;\leq\; c_2\,e^{-c_3 b_1 \log m} + c_5\,e^{-c_6 b_2 \log n} = o(m^{-1}) + o(n^{-1})$. Furthermore, in view of (23),

$$
P\big\{ \sup_{0\leq s\leq 1} \big|\mathbb{B}_{mn}(s)\big| \geq \sqrt{\log m} \big\} \;=\; 2\sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2(\sqrt{\log m})^2} \;=\; \mathcal{O}(m^{-2}).
$$

Thus, $T'_{mn}(i) = o(m^{-1}) + o(n^{-1}) + \mathcal{O}(m^{-2})$. Similarly, $T''_{mn}(i) = P\big\{ \sup_{0\leq s\leq 1} \big|\mathbb{B}_{mn}(s)\big| \geq \sqrt{\log m} \big\} = \mathcal{O}(m^{-2})$. Therefore, by (29),

$$
T_{mn}(i) \quad = \quad o(m^{-1}) + o(n^{-1}) + \mathcal{O}(m^{-2}). \qquad (31)
$$

To deal with the term $T_{mn}(ii)$ in (28), first observe that $\sup_{a\leq t\leq b} |f_n(t) - f(t)| \leq \sup_{a\leq t\leq b} \big|E[f_n(t)] - f(t)\big| + \sup_{a\leq t\leq b} \big|f_n(t) - E[f_n(t)]\big| = C_{12}\cdot h + \mu_K \sup_{x\in\mathbb{R}} |F_n(x) - F(X)| \cdot h^{-1}$, where one may take $C_{12} = \sup_{a\leq t\leq b} |f'(t)| \int |x| K(x) dx > 0$ (see, for example, Prakasa Rao (1983; page 47)). Therefore,

$$
\begin{aligned}
T_{mn}(ii) \quad &\leq \quad P\Big\{ \big[C_{12}\cdot h + \mu_K \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \cdot h^{-1} \big] \cdot \sup_{0\leq s\leq 1} \Big|\mathbb{B}_{mn}(s)\Big| \geq \frac{C_{11}\,\lambda^{1/2}h^{1/2}\varepsilon''_{mn,2}}{\ell_h\,\mu_K} \Big\} \\
&\leq \quad P\Big\{ \sup_{0\leq s\leq 1} \Big|\mathbb{B}_{mn}(s)\Big| \;\geq\; \frac{C_{11}\,\lambda^{1/2}h^{1/2}\varepsilon''_{mn,2}}{\ell_h\,\mu_K} \cdot \frac{1}{C_{12}\cdot h + \mu_K\sqrt{\frac{\log n}{n}}\cdot h^{-1}} \Big\} \\
&\quad + P\Big\{ \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \sqrt{\frac{\log n}{n}} \Big\} \qquad (32)
\end{aligned}
$$

Now, taking

$$
\varepsilon''_{mn,2} = \frac{\big[C_{12}\cdot h + \mu_K\sqrt{\frac{\log n}{n}}\cdot h^{-1} \big]\,\ell_h\,\mu_K\,\sqrt{\log n}}{C_{11}\,\lambda^{1/2}h^{1/2}}\,, \qquad (33)
$$

one finds

$$
T_{mn}(ii) \quad \leq \quad P\Big\{ \sup_{0\leq s\leq 1} \big|\mathbb{B}_{mn}(s)\big| > \sqrt{\log n} \Big\} + P\Big\{ \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \sqrt{(\log n)/n} \Big\}. \qquad (34)
$$

But, as before, $P\Big\{\sup_{0\le s\le 1}\big|\mathbb{B}_{mn}(s)\big| > \sqrt{\log n}\Big\} = \mathcal{O}(n^{-2})$. Furthermore, by the results of Dvoretzky et al. (1956) and Massart (1990),

$$P\Big\{\sup_{x\in\mathbb{R}}|F_n(x)-F(x)| > \sqrt{(\log n)/n}\Big\} \;\le\; 2\,n^{-2}.$$

This together with (34), (31), and (28) gives $S'_{n,1}(ii) = o(m^{-1}) + o(n^{-1}) + \mathcal{O}(m^{-2}) + \mathcal{O}(n^{-2})$. Combining this with (27) and (26), yields

$$S_{n,1}(ii) \;\le\; S'_{n,1}(ii) + S''_{n,1}(ii) \;=\; o(m^{-1}) + o(n^{-1}) + \mathcal{O}(m^{-2}) + \mathcal{O}(n^{-2}) \;=\; o(m^{-1}). \tag{35}$$

Therefore, in view of (35), (25), and (15), one finds

$$S_{n,1} \;\le\; o(m^{-1}) + o(n^{-1}) + \mathcal{O}(m^{-2}) + \mathcal{O}(n^{-2}) \;=\; o(m^{-1}), \tag{36}$$

where $S_{n,1}$ is as in (14). Finally, to deal with the term $S_{n,2}(x)$ in (14), let $\{\mathbb{B}(s),\ 0\le s\le 1\}$ be a Brownian bridge and note that for each $n = 1,2,\ldots$ and $m = 1,2,\ldots,$

$$\Big\{\frac{1}{\sqrt{f(t)}}\int K((t-s)/h)\,d\,\mathbb{B}_{mn}(F(s)),\ a\le t\le b\Big\}$$
$$\stackrel{\mathcal{D}}{=} \Big\{\frac{1}{\sqrt{f(t)}}\int K((t-s)/h)\,d\,\mathbb{B}(F(s)),\ a\le t\le b\Big\}. \tag{37}$$

Konakov and Piterbarg (1984) studied the process $(f(t))^{-1/2}\int K((t-s)/h)\,d\,\mathbb{B}(F(s)),\ \ a\le t\le b$. Their results show that, under assumptions (A), (B), and (C), for the random variable

$$\widetilde{M}_n^0 = \ell_h\,\lambda^{-1/2}h^{-1/2}\sup_{a\le t\le b}\left|\big(1/\sqrt{f(t)}\,\big)\int K((t-s)/h)\,d\,\mathbb{B}(F(s))\right| - \ell_h^2\,, \tag{38}$$

there exists a constant $\beta > 0$ such that $P\big\{\widetilde{M}_n^0 \le x\big\} = \exp\big\{-2\exp\big\{-x-\frac{x^2}{2\ell_h^2}\big\}\big\} + \mathcal{O}(n^{-\beta})$, uniformly in x. Now, let $G(u) = \exp\big\{-2\exp\big\{-u-\frac{u^2}{2\ell_h^2}\big\}\big\}$ and observe that by virtue of (37)

$$S_{n,2}(x) \;=\; P\big\{\big|\widetilde{M}_n^0 - x\big| < \varepsilon_{mn}\big\} \;=\; G(x+\varepsilon_{mn}) - G(x-\varepsilon_{mn}) + \mathcal{O}(n^{-\beta}),$$

uniformly in x. Since G is differentiable (everywhere), the mean value theorem for integrals yields

$$G(x+\varepsilon_{mn}) - G(x-\varepsilon_{mn}) \;=\; \int_{x-\varepsilon_{mn}}^{x+\varepsilon_{mn}} G'(u)\,du \;=\; 2\varepsilon_{mn}\cdot G'(c)\,,$$

for some $c\in(x-\varepsilon_{mn},\,x+\varepsilon_{mn})$. Therefore,

$$S_{n,2}(x) = \mathcal{O}\big(n^{-\lambda}(\log n)^{3/2}\big) + \mathcal{O}(n^{-\beta}),\ \text{uniformly in } x, \tag{39}$$

$\lambda > 0$, which follows from the fact that G' is bounded and the fact that $\varepsilon_{mn} = \varepsilon'_{mn} + \varepsilon''_{mn,1} + \varepsilon''_{mn,2} = \mathcal{O}\big(\ell_h\log m/\sqrt{mh}\big) + \mathcal{O}\big(\ell_h\log m/\sqrt{mh}\big) + \mathcal{O}\big(\ell_h\log n/\sqrt{nh^3}\big) = \mathcal{O}\big(n^{-\lambda}\,(\log n)^{3/2}\big), \lambda > 0$, because $\ell_h = \mathcal{O}(\sqrt{\log n})$. Putting (39) together with (14), (36), and (37), we find $\big|P\big\{\widehat{M}_{mn}\le x\big\} - P\big\{\widetilde{M}_{mn}\le x\big\}\big| = \mathcal{O}\big(n^{-\lambda}\,(\log n)^{3/2}\big) + \mathcal{O}(n^{-\beta})$, uniformly in x, where $\widetilde{M}_{mn}$ is as in (13). Since $P\big\{\widetilde{M}_{mn}\le x\big\} = P\big\{\widetilde{M}_n^0\le x\big\}$ for all x, m, and n, this means

$$\big|P\big\{\widehat{M}_{mn}\le x\big\} - P\big\{\widetilde{M}_n^0\le x\big\}\big| \;=\; \mathcal{O}\big(n^{-\lambda}\,(\log n)^{3/2}\big) + \mathcal{O}(n^{-\beta})\,, \tag{40}$$

uniformly in x. Similarly, one can show

$$\big|P\{\widehat{M}_n \le x\} - P\{\widetilde{M}_n^0 \le x\}\big| = \mathcal{O}\big(n^{-\lambda}\,(\log n)^{3/2}\big) + \mathcal{O}(n^{-\beta})\,, \tag{41}$$

uniformly in x, where $\widehat{M}_n$ is as in (9). The proof of (41) is similar to that of (40); the key steps of the proof are as follows. Let $\{\mathbb{B}_n(t),\ 0 \le t \le 1\}$, $n = 1, 2, \ldots$, be the sequence of Brownian bridges used by Komlós, et al. (1975) in the approximation of empirical processes, and put

$$\widetilde{M}_n \;=\; \ell_h\, \lambda^{-1/2} h^{-1/2} \sup_{a\le t\le b} \left| \big(1/\sqrt{f(t)}\,\big) \int K((t-s)/h)\, d\,\mathbb{B}_n(F(s)) \right| - \ell_h^2\,.$$

Now, in view of Lemma 2 and arguments similar to those that led to (14) and (15), for every $\gamma_n > 0$, $\gamma_n' > 0$, and $\gamma_n'' > 0$, with $\gamma_n' + \gamma_n'' = \gamma_n$, one obtains

$$\begin{aligned}
&\big|P\{\widehat{M}_n \le x\} - P\{\widetilde{M}_n \le x\}\big| \\
&\le P\Big\{\sup_{a\le t\le b} \frac{1}{\sqrt{f_n(t)}} \Big| \int K((t-s)/h)\, d\beta_n(s) - \int K((t-s)/h)\, d\,\mathbb{B}_n(F(s)) \Big| \ge \frac{\lambda^{1/2} h^{1/2} \gamma_n'}{\ell_h} \Big\} \\
&\quad + P\Big\{\sup_{a\le t\le b} \Big| \Big[\frac{1}{\sqrt{f_n(t)}} - \frac{1}{\sqrt{f(t)}}\Big] \int K((t-s)/h)\, d\,\mathbb{B}_n(F(s)) \Big| \ge \frac{\lambda^{1/2} h^{1/2} \gamma_n''}{\ell_h} \Big\} \\
&\quad + P\Big\{\Big| \ell_h \lambda^{-1/2} h^{-1/2} \sup_{a\le t\le b} \Big| \frac{1}{\sqrt{f(t)}} \int K((t-s)/h)\, d\,\mathbb{B}_n(F(s)) \Big| - x - \ell_h^2 \Big| < \gamma_n \Big\} \\
&=: U_{n1} + U_{n2} + U_n(x)\,.
\end{aligned} \tag{42}$$

The term U_{n1} can be bounded based on the arguments similar to those that led to (19), (21), and (25). More specifically, we can find a positive constant C_{16} such that, for the choice

$$\gamma_n' = \ell_h\, \mu_K \log n \cdot C_{16} \big/ \big[(f_0/2)^{1/2} \lambda^{1/2} h^{1/2} n^{1/2}\big] \tag{43}$$

one has

$$\begin{aligned}
U_{n1} \;&\le\; P\Big\{ \frac{\mu_K}{\sqrt{f_0/2}} \sup_{v\in\mathbb{R}} \big|\beta_n(v) - \mathbb{B}_n(F(v))\big| \ge \frac{\lambda^{1/2} h^{1/2} \gamma_n'}{\ell_h} \Big\} + P\Big\{ \sup_{a\le t\le b} \big|f_n(t) - f(t)\big| > f_{\max} - \frac{f_0}{2} \Big\} \\
&=\; o(n^{-1}) \,+\, \mathcal{O}(\exp(-c_{15} n h^2)) \;=\; o(n^{-1}),
\end{aligned} \tag{44}$$

where $\beta_n(\cdot)$ is as in (12). The term U_{n2} in (42) can be bounded as in (34). More specifically, choosing

$$\gamma_n'' = \Big(C_{12} \cdot h + \mu_K \sqrt{(\log n)/n} \cdot h^{-1}\Big)\, \ell_h\, \mu_K \sqrt{\log n} \,\big/\, \big(C_{17}\, \lambda^{1/2} h^{1/2}\big)\,, \tag{45}$$

where $0 < C_{12} = \sup_{a\le t\le b} |f'(t)| \int |x| K(x) dx < \infty$ and $C_{17} = \sqrt{f_0/2}\, f_0 + (f_0/2)\sqrt{f_0}$, one finds

$$\begin{aligned}
U_{n2} \;&\le\; P\Big\{ \sup_{a\le t\le b} |f_n(t) - f(t)| \cdot \sup_{0\le s\le 1} \big|\mathbb{B}_n(s)\big| \;\ge\; C_{17}\, \lambda^{1/2} h^{1/2} \gamma_n'' / (\ell_h\, \mu_K) \Big\} \\
&\le\; P\Big\{ \sup_{0\le s\le 1} \big|\mathbb{B}_n(s)\big| > \sqrt{\log n} \Big\} + P\Big\{ \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| > \sqrt{n^{-1}\log n} \Big\} = \mathcal{O}(n^{-2}).
\end{aligned} \tag{46}$$

Finally, the arguments that led to (39) and the fact that $\gamma_n = \gamma_n' + \gamma_n'' = \mathcal{O}\big(\ell_h \log n/\sqrt{nh}\big) + \mathcal{O}\big(\ell_h \log n/\sqrt{nh^3}\big) = \mathcal{O}\big(n^{-\lambda}\,(\log n)^{3/2}\big)$, yields $U_n(x) = \mathcal{O}\big(n^{-\lambda}\,(\log n)^{3/2}\big) + \mathcal{O}(n^{-\beta})$, uniformly in x.

Now, (41) follows because $P\{\widetilde{M}_n \le x\} = P\{\widetilde{M}_n^0 \le x\}$ for all x and n. Part (i) of Theorem 2 now follows from (40) and (41).

Proof of part (ii)

The proof of part (ii) is similar to (and in fact easier than) that of part (i) and will not be given. □

3 Numerical examples

3.1 The choice of m in practice

We start by discussing a qualitative approach for the choice of m which is in the spirit of the method of *successive differences*, proposed by Bickel and Sakov (2008). Let M_{mn} and $\widehat{M}_{mn}$ be, respectively, the non-studentized and studentized bootstrap statistics in (10). Also, for each fixed m, let $M_{mn,1}, \ldots, M_{mn,B}$ be B copies of M_{mn} based on B bootstrap sample of size m (B is typically a large number such as 1000). Similarly, let $\widehat{M}_{mn,1}, \ldots, \widehat{M}_{mn,B}$ be B copies of M_{mn}, and define

$$L^*_{m,n}(x) \;=\; B^{-1}\sum_{b=1}^{B} I\{M_{mn,b} \le x\} \quad \text{and} \quad \widehat{L}^*_{m,n}(x) \;=\; B^{-1}\sum_{b=1}^{B} I\left\{\widehat{M}_{mn,b} \le x\right\}.$$

Then the method of *successive differences* chooses the smallest m_j that approximately minimizes

$$d_j = \sup_x \left| L^*_{m_j,n}(x) - L^*_{m_{j+1},n}(x) \right|, \tag{47}$$

where m_j, $j \ge 1$, is an increasing sequence of positive integers. In the case of the studentized statistic $\widehat{M}_{mn}$, m_j is chosen to be the smallest value that approximately minimizes

$$\widehat{d}_j = \sup_x \left| \widehat{L}^*_{m_j,n}(x) - \widehat{L}^*_{m_{j+1},n}(x) \right|. \tag{48}$$

In practice, the supremum is approximated by taking the maximum over a grid. For example, (47) can be approximated by taking the maximum over a grid of equally-spaced values of M_{mn} in the range $[\min_{1\le b\le B} M_{mn,b}\,,\, \max_{1\le b\le B} M_{mn,b}]$ for some initial value of m. Similarly, in the case of (48), m_j is chosen to minimize $\widehat{d}_j$ over a grid of values of $\widehat{M}_{mn}$ in $\left[\min_{1\le b\le B} \widehat{M}_{mn,b}\,,\, \max_{1\le b\le B} \widehat{M}_{mn,b}\right]$. We illustrate this approach in the next subsection.

3.2 Simulated data and the choice of m

To see how the above method works, here we carry out a simulation study to assess the performance of the bootstrap approximation when $m << n$ and n is large. We consider a random sample of size $n = 10^6$ drawn from the mixture distribution $f(x) = 0.4\phi_1(x) + 0.6\phi_2(x)$, where ϕ_1 is the pdf of the $N(\mu = -3,\, \sigma = 1)$ distribution and ϕ_2 is that of $N(\mu{=}1,\, \sigma{=}2)$. Next, $B{=}1000$ copies of the bootstrap statistic M_{mn} and its "studentized" counterpart $\widehat{M}_{mn}$ in (10) were constructed, which were then used to find the successive difference in (47) and (48) for different values of m. A plot of m against the successive differences (47) appears in the top panel of Figure 1, whereas that of (48) is given in the lower panel. Figure 1 shows a value of m around 1400 may be sufficient in the case of statistic M_{mn}. This number is lower (around 900) in the case of the studentized statistics $\widehat{M}_{mn}$, as given by (10). In this example, the bandwidth h was estimated using the method of Sheather and Jones (1991) subject to the conditions of Theorem 2.

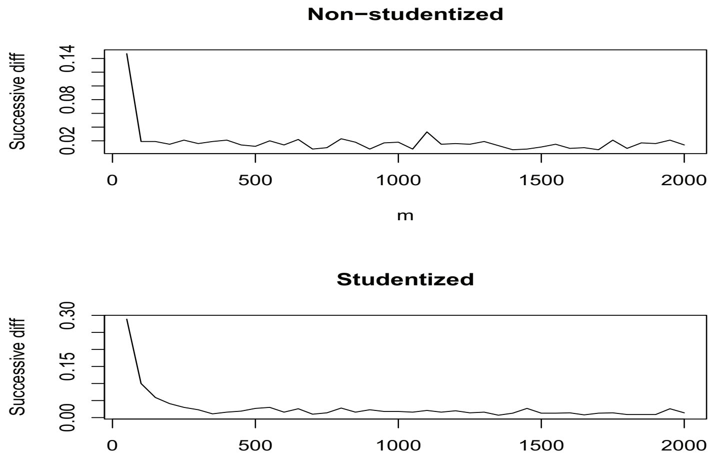

Figure 1: Plots of different bootstrap sample sizes m against the *successive differences*.

Next, let $U = \exp(-2\exp\{M_n\})$ and $\widehat{U} = \exp\left(-2\exp\left\{\widehat{M_n}\right\}\right)$, where M_n and $\widehat{M_n}$ are as in (9), and observe that when n is large, then by Theorem 1 the random variables U and $\widehat{U}$ should be approximately Unif (0,1) random variables. Similarly, by Theorem 2, the random variables $T = B^{-1}\sum_{b=1}^{B} I\{M_{mn,b} \leq M_n\}$ and $\widehat{T} = B^{-1}\sum_{b=1}^{B} I\{\widehat{M}_{mn,b} \leq \widehat{M_n}\}$ should be approximately Unif(0,1). Now, drawing 300 independent random samples, each of size $n = 10^6$ from $f(x)$, we obtain $(U_1, \ldots, U_{300})$, $(\widehat{U}_1, \ldots, \widehat{U}_{300})$, $(T_1, \ldots, T_{300})$, and $(\widehat{T}_1, \ldots, \widehat{T}_{300})$. The top panel of Figure 2 illustrates the plot of the empirical CDF of the U_i's in (a), that of T_i's in (b), $\widehat{U}_i$'s in (c), and $\widehat{T}_i$'s in (d). We have also included the 45-degree line which is the theoretical CDF of the Unif(0,1) distribution. The fact that plots (b) and (d) are much closer to the 45-degree line (as compared to plots (a)

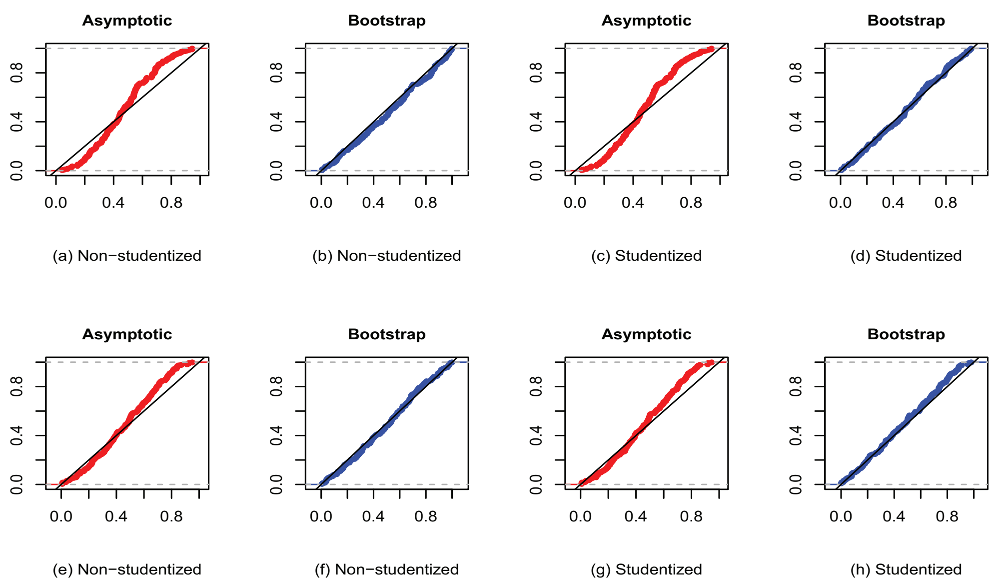

Figure 2: Plots of the empirical cdf's. Top panel corresponds to the density of a mixture of two normals: plot (a) is for $U_1, \cdots, U_{300}$, plot (b) for $T_1, \cdots, T_{300}$, plot (c) for $\widehat{U}_1, \ldots, \widehat{U}_{300}$, and plot (d) corresponds to $\widehat{T}_1, \ldots, \widehat{T}_{300}$. Plots (e)–(h) correspond to the density of the mixture of a normal and a uniform distributions.

and (c)) shows the superiority of the bootstrap. Plots (e) – (h) produce the same results when the underlying density is the mixture $f(x) = 0.4\varphi_1(x) + 0.6\varphi_2(x)$, where φ_1 is the density of a standard normal and φ_2 is the density of a Unif$(-3, 2)$ distribution. Once again, we see the superiority of the bootstrap for re-sample sizes that are substantially smaller than n.

3.3 A real data example and the choice of m

Here we consider a real data example, with more than 500,000 entries, involving the unit prices of items in a large online retail data set. A full description of this data set is available at the UCI repository of machine learning data sets: `https://archive.ics.uci.edu/ml/datasets/online+retail`. Chen et al. (2012) have also studied this data set. Here, we consider the application of successive differences in order to be able to approximate the smallest re-sample size m that can be used to construct the same bootstrap statistics. Since some of the price entries are reported to be zero or take negative values, it was decided to remove such entries first. The remaining number of cases is still close to 500,000. Figure 3 shows the plot of m versus successive differences. Here, a value of m

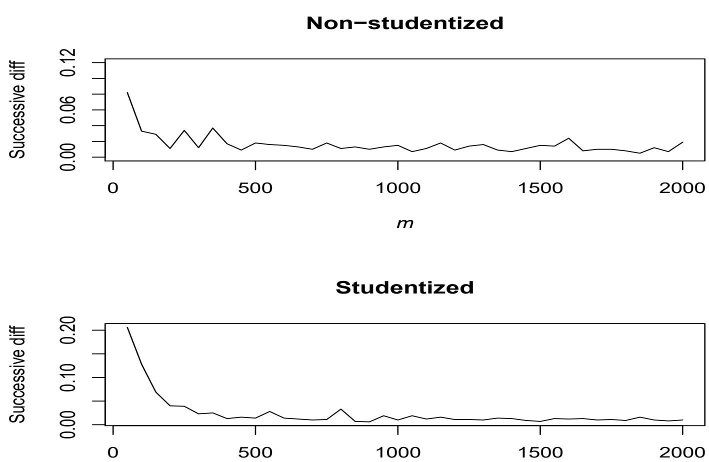

Figure 3: Plots of different bootstrap sample sizes m against the *successive differences*.

about 1000 for the non-studentized and 1500 for the studentized statistics might be sufficient.

Concluding Remarks. In this paper, we have studied the bootstrap approximation of the distribution of the maximal deviations of kernel density estimators in big-data contexts. To appreciate the difficulties involved, we note that when the sample size n is very large and when the computation of statistics of interest is quite involved, then the computational cost associated with the use of bootstrap samples of the same size as the original data can be formidable. Of course, large data sizes can be blessings when invoking asymptotic results, but they can also be hindering in re-sampling methods such as the bootstrap. The proposed *virtual bootstrap* provides a solution to reduce this computational cost while still retaining the benefits of bootstrap methodology.

Acknowledgements
This work was supported by the NSF under Grant DMS-1916161 of Majid Mojirsheibani.

References

Alvarez-Andrade, S., Bouzebda, S., 2013. Strong approximations for weighted bootstrap of empirical and quantile processes with applications. Statist. Methodol. 11, 36-52.

Al-Sharadqah, A., Mojirsheibani, M., Pouliot, W., 2020. On the performance of weighted bootstrapped kernel deconvolution density estimators. Statist. Papers 61, 1773-1798.

Bickel, P., Götze, F., van Zwet, W., 1997. Resampling fewer than n observations: gains, losses and remedies for losses. Statistica Sinica, 7, 1-31.

Bickel, P., Sakov, A., 2008. On the choice of m in m out of n bootstrap and confidence bounds for extrema. Statistica Sinica, 18, 967-985.

Bickel, P., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Ann. Statist. 1, 1071-1095.

Chen, D., Sain, S.L., Guo, K., (2012) Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining, Journal of Database Marketing and Customer Strategy Management, Vol. 19, No. 3.

Csörgő, M., Horváth, L., Kokoszka, P., 1999. Approximation for bootstrapped empirical processes. Proc. Amer. Math. Soc. 128, 2457-2464.

Csörgő, M., Horváth, L., 1993. Weighted Approximations in Probability and Statistics. Wiley.

Dvoretzky, A., Kiefer, J., Wolfowitz, J., 1956. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist. 27, 642-669.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1-26.

Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent random variables, and the sample distribution function. Z. Wahrscheinlich. Verwandte Geb. 32, 111-131.

Konakov, V.D., Piterbarg, V.I., 1984. On the convergence rate of maximal deviation distribution for kernel regression estimates. J. Multivariate Anal. 15, 279-294.

Massar, P. 1999. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. Ann. Probab. 18, 1269-1283.

Mojirsheibani, M., 2012. A weighted bootstrap approximation of the maximal deviation of kernel density estimates over general compact sets. J. Multivariate Anal. 112, 230-241.

Muminov, M.S., 2011. On the limit distribution of the maximum deviation of the empirical distribution density and the regression function. I. Theory Probab. Appl. 55, 509-517.

Muminov, M.S., 2012. On the limit distribution of the maximum deviation of the empirical distribution density and the regression function. II. Theory Probab. Appl. 56, 155-166.

Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-1076.

Prakasa Rao, B., 1983. Nonparametric Functional Estimation. Academic Press, Orlando.

Rio, E., 1994. Local invariance principles and their applications to density estimation. Probab. Theory and Related Fields. 98, 21-45.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27, 832-837.

Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. J. Royal Statist. Soc. Ser. B, 53, 683-690.