PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 122, Number 4, December 1994

BOOTSTRAP SAMPLE SIZE IN NONREGULAR CASES

JUN SHAO

(Communicated by Wei Y. Loh)

ABSTRACT. We study the bootstrap estimator of the sampling distribution of a given statistic in some nonregular cases where the given statistic is nonsmooth or not-so-smooth. It is found that the ordinary bootstrap, based on a bootstrap sample of the same size as the original data set, produces an inconsistent bootstrap estimator. On the other hand, when we draw a bootstrap sample of a smaller size with the ratio of the size of the bootstrap sample over the size of the original data set tending to zero, the resulting bootstrap estimator is consistent. Examples of these nonregular cases are given, including the cases of functions of means with null first-order derivatives, differentiable statistical functionals with null influence function, nondifferentiable functions of means, and estimators based on some test results.

1. INTRODUCTION

The bootstrap method [7] is shown to be successful in many situations. In fact, it is better than some other asymptotic methods, such as the traditional normal approximation and the Edgeworth expansion. See the results established by Abramovitch and Singh in [1], Beran in [4], Efron in [9], and Hall in [10, 11]. However, there are some counterexamples that show the bootstrap produces wrong solutions, i.e., it provides some inconsistent estimators [5, 3, 14, 2, 13].

We focus on the case where the data $X_1, \dots, X_n$ are independent and identically distributed (i.i.d.) samples from a k-dimensional population distribution F and the problem of estimating the distribution

$$H_n(x) = P\{R_n \le x\}, \tag{1.1}$$

where $R_n = R_n(T_n, F)$ is a real-valued functional of F and $T_n = T_n(X_1, \dots, X_n)$, a statistic of interest. Let $X_1^*, \dots, X_n^*$ be "bootstrap" samples i.i.d. from F_n, the empirical distribution based on $X_1, \dots, X_n$, $T_n^* = T_n(X_1^*, \dots, X_n^*)$, and $R_n^* = R_n(T_n^*, F_n)$. A bootstrap estimator of H_n is

$$\widehat{H}_n(x) = P_*\{R_n^* \le x\}, \tag{1.2}$$

Received by the editors March 3, 1993.
1991 *Mathematics Subject Classification*. Primary 62G09; Secondary 62E20.
Key words and phrases. Bootstrap sample size, consistency, differentiable functionals, estimators based on tests, functions of means.
The research was supported by the Natural Sciences and Engineering Research Council of Canada.

where for given $X_1, \ldots, X_n$, P_* is the conditional probability with respect to the random generator of bootstrap samples. Since the bootstrap samples are generated from F_n, this method is called the nonparametric bootstrap.

Although the bootstrap estimator $\widehat{H}_n$ is consistent for many commonly used statistics T_n, there are some cases where $\widehat{H}_n$ is inconsistent. The main reasons for the inconsistency of $\widehat{H}_n$ are:

(1) The bootstrap is sensitive to the tail behavior of F. Some moment conditions are needed for the consistency of $\widehat{H}_n$.

(2) The consistency of $\widehat{H}_n$ requires that the given statistic T_n be smooth, i.e., T_n can be well approximated by a linear statistic.

The moment and smoothness conditions required for the consistency of $\widehat{H}_n$ are usually assumed as regularity conditions in the literature. Situations of lack of these conditions are then called nonregular cases. Examples of inconsistency of $\widehat{H}_n$ caused by the lack of moment conditions are given by Bickel and Freedman [5], Babu [3], and Athreya [2]. In this paper we concentrate on the other nonregular case: T_n is nonsmooth or not-so-smooth so that T_n is not asymptotically normal. We provide some examples of inconsistency of $\widehat{H}_n$, and more importantly, we show that the inconsistency caused by nonsmoothness can be rectified by changing the bootstrap sample size n (the number of bootstrap samples taken from F_n) to m with $m = m_n \to \infty$ and $m_n/n \to 0$ (or $m_n \log\log n/n \to 0$) as $n \to \infty$. We will also explain why this change produces a consistent bootstrap estimator.

2. Functions of means with null first-order differential

Let $\mu = EX_1$; $\theta = g(\mu)$, where g is a function from $\mathbb{R}^k$ to $\mathbb{R}$; $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ be the sample mean; and $T_n = g(\overline{X}_n)$. Under the conditions that $\operatorname{var}(X_1) = \Sigma < \infty$ and g is first-order continuously differentiable at μ,

$$\sup_x \left|\{\sqrt{n}(T_n^* - T_n) \le x\} - P\{\sqrt{n}(T_n - \theta) \le x\}\right| \to 0 \quad \text{a.s.}, \tag{2.1}$$

where $T_n^* = g(\overline{X}_n^*)$ and $\overline{X}_n^* = n^{-1}\sum_{i=1}^n X_i^*$.

Consider the situation where g is second-order continuously differentiable at μ with $\nabla^2 g(\mu) \neq 0$ but $\nabla g(\mu) = 0$. Using the Taylor expansion and $\nabla g(\mu) = 0$, we obtain that

$$T_n - \theta = \tfrac{1}{2}(\overline{X}_n - \mu)'\nabla^2 g(\mu)(\overline{X}_n - \mu) + o_p(n^{-1}). \tag{2.2}$$

This implies

$$n(T_n - \theta) \to_d \tfrac{1}{2} Z_\Sigma' \nabla^2 g(\mu) Z_\Sigma, \tag{2.3}$$

where Z_Σ is a random k-vector having normal distribution with mean 0 and covariance matrix Σ. From (2.3), $\sqrt{n}(T_n - \theta) \to_p 0$, and, therefore, result (2.1) is not useful when $\nabla g(\mu) = 0$ and we need to consider the bootstrap estimator of the distribution of $n(T_n - \theta)$ in this case.

Let $R_n = n(T_n - \theta)$, $R_n^* = n(T_n^* - T_n)$, and H_n and $\widehat{H}_n$ be given by (1.1) and (1.2), respectively. Buba [3] pointed out that $\widehat{H}_n$ is inconsistent in this case. Similar to (2.2),

$$\begin{aligned} T_n^* - T_n = {} & \nabla g(\overline{X}_n)'(\overline{X}_n^* - \overline{X}_n) \\ & + \tfrac{1}{2}(\overline{X}_n^* - \overline{X}_n)'\nabla^2 g(\overline{X}_n)(\overline{X}_n^* - \overline{X}_n) + o_p(n^{-1}) \quad \text{a.s.} \end{aligned} \tag{2.4}$$

By the continuity of $\nabla^2 g$ and Theorem 2.1 in [5], for almost all given sequences $X_1, X_2, \dots,$

$$\tfrac{n}{2}(\overline{X}_n^* - \overline{X}_n)'\nabla^2 g(\overline{X}_n)(\overline{X}_n^* - \overline{X}_n) \to_d \tfrac{1}{2} Z_\Sigma' \nabla^2 g(\mu) Z_\Sigma . \tag{2.5}$$

From $\nabla g(\mu) = 0$,

$$\sqrt{n}\nabla g(\overline{X}_n) = \sqrt{n}\nabla^2 g(\mu)(\overline{X}_n - \mu) + o_p(1) \to_d \nabla^2 g(\mu) Z_\Sigma . \tag{2.6}$$

Hence, for almost all given $X_1, X_2, \dots,$ the conditional distribution of $n\nabla g(\overline{X}_n)'(\overline{X}_n^* - \overline{X}_n)$ does not have a limit. It follows from (2.4) and (2.5) that for almost all given $X_1, X_2, \dots,$ the conditional distribution of $n(T_n^* - T_n)$ does not have a limit. Therefore, $\widehat{H}_n$ is inconsistent as an estimator of H_n.

This inconsistency is caused by an inherent problem of the bootstrap: The bootstrap samples are drawn from F_n which is not exactly F. The effect of this problem is inappreciable in a regular case (i.e., T_n can be well approximated by a linear statistic) but leads to inconsistency in a nonregular case (by (2.2), T_n is well approximated by a quadratic statistic but not a linear statistic). The symptom of this problem in the present case is that $\nabla g(\overline{X}_n)$ is not necessarily equal to zero when $\nabla g(\mu) = 0$. As a result, the expansion in (2.4), compared with the expansion in (2.2), has an extra nonzero term $\nabla g(\overline{X}_n)'(\overline{X}_n^* - \overline{X}_n)$ which does not converge to zero fast enough, and, therefore, $\widehat{H}_n$ cannot mimic H_n.

When we take bootstrap samples, it is not necessary that we should always take n samples. Let m be an integer, $X_1^*, \dots, X_m^*$ be i.i.d. from F_n, $R_{n,m}^* = R_m(T_m^*, F_n)$, and $T_m^* = T_m(X_1^*, \dots, X_m^*)$ be the bootstrap analogues of R_n and T_n, respectively, based on m bootstrap samples. Then we may use the following bootstrap estimator of H_n:

$$\widehat{H}_{n,m}(x) = P_*\{R_{n,m}^* \le x\} . \tag{2.7}$$

Bickel and Freedman [5] actually studied this bootstrap estimator with m being a function of n or with m varying free with n. There is no apparent reason why we should always use $m = n$, but $m = n$ $(\widehat{H}_{n,m} = \widehat{H}_n)$ is customarily used and works well for regular cases. However, we will show that allowing a bootstrap sample size m different from n gives us more freedom for rectifying the inconsistency of the bootstrap estimator in nonregular cases.

If we could take the sample size $n = \infty$, then $\widehat{H}_{n,m}$ is consistent $(m \to \infty)$, since $X_1^*, \dots, X_m^*$ can be viewed as samples from F when $n = \infty$. Of course, taking $n = \infty$ is impractical. However, we may select $m = m_n$ so that F_n converges to F faster than its bootstrap analogue F_m^*, the empirical distribution based on $X_1^*, \dots, X_m^*$, and achieve the same effect as taking $n = \infty$ (or generating the bootstrap samples from F). Note that $\|F_n - F\|_\infty = O_p(n^{-1/2})$ or $O(\sqrt{\log\log n/n})$ a.s. (see, e.g., [15, p. 62]), where $\|\cdot\|_\infty$ is the sup-norm, and $\|F_m^* - F_n\|_\infty = O_p(m^{-1/2})$ [5]. Therefore, for the weak consistency of $\widehat{H}_{n,m}$ we may let $m = m_n$ with $m_n \to \infty$ and $m_n/n \to 0$ as $n \to \infty$; for the strong consistency of $\widehat{H}_{n,m}$ we may let $m = m_n$ with $m_n \to \infty$ and $m_n \log\log n/n \to 0$. We now show that this change of bootstrap sample size does provide consistent bootstrap estimators in nonregular cases where $\widehat{H}_n$ is inconsistent.

Theorem 1. *Suppose that* $\operatorname{var}(X_1) < \infty$ *and* g *is second-order continuously differentiable at* μ *with* $\nabla^2 g(\mu) \neq 0$ *and* $\nabla g(\mu) = 0$.

(i) *If* $m = m_n \to \infty$ *and* $m_n/n \to 0$, *then* $\widehat{H}_{n,m}$ *is weakly consistent, i.e.,*

$$\sup_x |\widehat{H}_{n,m}(x) - H_n(x)| \to_p 0. \tag{2.8}$$

(ii) *If* $m = m_n \to \infty$ *and* $m_n \log\log n/n \to 0$, *then* $\widehat{H}_{n,m}$ *is strongly consistent, i.e.,*

$$\sup_x |\widehat{H}_{n,m}(x) - H_n(x)| \to 0 \quad \textit{a.s.} \tag{2.9}$$

Proof. By Taylor expansion,

$$\begin{aligned} T_m^* - T_n &= \nabla g(\overline{X}_n)'(\overline{X}_m^* - \overline{X}_n) \\ &\quad + \tfrac{1}{2}(\overline{X}_m^* - \overline{X}_n)'\nabla^2 g(\overline{X}_n)(\overline{X}_m^* - \overline{X}_n) + o_p(m^{-1}) \quad \text{a.s.} \end{aligned} \tag{2.10}$$

By Theorem 2.1 in [5], for almost all given sequences $X_1, X_2, \ldots,$

$$\tfrac{m}{2}(\overline{X}_m^* - \overline{X}_n)'\nabla^2 g(\overline{X}_n)(\overline{X}_m^* - \overline{X}_n) \to_d \tfrac{1}{2} Z_\Sigma' \nabla^2 g(\mu) Z_\Sigma. \tag{2.11}$$

The expansion in (2.10) still has the nonzero term $\nabla g(\overline{X}_n)'(\overline{X}_m' - \overline{X}_n)$, but by (2.6), it is of the order $o_p(m^{-1})$ in probability if $m/n \to 0$ and $o_p(m^{-1})$ a.s. if $m \log\log n/n \to 0$. This proves that

$$T_m^* - T_n = \tfrac{1}{2}(\overline{X}_m^* - \overline{X}_n)'\nabla^2 g(\overline{X}_n)(\overline{X}_m^* - \overline{X}_n) + o_p(m^{-1}) \tag{2.12}$$

in probability or a.s., as if we could take $n = \infty$ or $\nabla g(\overline{X}_n)$ could be treated as zero. Then, results (2.8) and (2.9) follow from (2.11) and (2.12). □

3. Functionals with null first-order differential

The result in Theorem 1 can be extended to the general case where $T_n = T(F_n)$ with a second-order differentiable functional T having null first-order differential.

Definition. Let T be a functional on a convex class of distributions containing F and all degenerate distributions. T is said to be first-order Fréchet differentiable at F if there is a function $\phi_F(x)$ on $\mathbb{R}^k$ such that $\int \phi_F(x)\,dF(x) = 0$ and for any sequence $\{G_n\}$ satisfying $\|G_n - F\|_\infty \to 0$,

$$T(G_n) - T(F) - \int \phi_F(x)\,dG_n(x) = o(\|G_n - F\|_\infty).$$

T is said to be second-order Frèchet differentiable at F if there is a function $\phi_F(x)$ on $\mathbb{R}^k$ and a function $\psi_F(x, y)$ on $\mathbb{R}^{2k}$ such that $\int \phi_F(x)\,dF(x) = 0$, $\psi_F(x, y) = \psi_F(y, x)$, $\int \psi_F(x, y)\,dF(x) = 0$ for all y, and for any sequence $\{G_n\}$ satisfying $\|G_n - F\|_\infty \to 0$,

$$T(G_n) - T(F) - \int \phi_F(x)\,dG_n(x) - \iint \psi_F(x, y)\,dG_n(x)\,dG_n(y) = o(\|G_n - F\|_\infty^2).$$

The function ϕ_F in the above definition is the influence function of T [12]. Similar to $\nabla g(\mu) = 0$ and $\nabla^2 g(\mu) \neq 0$ in the last section, we consider the case of

$$\phi_F \equiv 0 \quad \text{and} \quad \psi_F \not\equiv 0. \tag{3.1}$$

Examples of $T_n = T(F_n)$ having zero influence function are some "goodness of fit" test statistics under null hypothesis. The following is an example.

Consider the test problem

$$H_0 : F \equiv F_0 \quad \text{versus} \quad H_1 : F \not\equiv F_0$$

with a given distribution F_0 and the statistic generated by the functional

$$T(G) = \int w_{F_0}(x)[G(x) - F_0(x)]^2 \, dF_0(x), \tag{3.2}$$

where $w_{F_0}(x)$ is a weight function depending on the known distribution F_0 and satisfying $\int w_{F_0}(x) \, dF_0(x) < \infty$. When $w_{F_0}(x) = 1$ for all x, $T(F_n)$ is the Cramér-von Mises test statistic.

It can be shown that T in (3.2) is first-order Fréchet differentiable at F with the influence function

$$\phi_F(x) = 2 \int w_{F_0}(y)[I\{x \le y\} - F(y)][F(y) - F_0(y)] \, dF_0(y), \tag{3.3}$$

where $I\{A\}$ is the indicator of the set A, and T is also second-order Fréchet differentiable at F with

$$\psi_F(x, y) = \int w_{F_0}(t)[I\{x \le t\} - F(t)][I\{y \le t\} - F(t)] \, dF_0(t). \tag{3.4}$$

Under the null hypothesis H_0, $F \equiv F_0$ and (3.1) holds. From Theorem 6.4.1B in [15], under H_0,

$$n[T(F_n) - T(F)] \to_d W,$$

which is a weighted sum of independent chi-square random variables.

Let $R_n = n[T(F_n) - T(F)]$, $R_n^* = n[T(F_n^*) - T(F_n)]$, H_n and $\widehat{H}_n$ be given by (1.1) and (1.2), respectively, and ϕ_{F_n} and ψ_{F_n} be given by (3.3) and (3.4), respectively, with F replaced by F_n. Then for T given by (3.2),

$$T(F_n^*) - T(F_n) = \int \phi_{F_n}(x) \, dF_n^*(x) + \iint \psi_{F_n}(x, y) \, dF_n^*(x) \, dF_n^*(y).$$

It can be shown that for almost all given sequences $X_1, X_2, \ldots$, the conditional distribution of $n \iint \psi_{F_n}(x, y) \, dF_n^*(x) \, dF_n^*(y)$ converges weakly to that of W, but the conditional distribution of

$$n \int \phi_{F_n}(x) \, dF_n^*(x) = 2n \int w_{F_0}(x)(F_n^* - F_n)(x)(F_n - F_0)(x) \, dF_0(x)$$

does not have a weak limit. This implies the inconsistency of the bootstrap estimator $\widehat{H}_n$.

The problem here is very similar to that in the last section: although $\phi_F \equiv 0$, $\phi_{F_n} \not\equiv 0$.

On the other hand, we have the following general result that shows the consistency of the bootstrap estimator $\widehat{H}_{n,m}$ defined in (2.7) with $R_{n,m}^* = m[T(F_m^*) - T(F_n)]$, where F_m^* is the empirical distribution based on a bootstrap of size m generated from F_n.

Theorem 2. *Suppose that T is second-order Fréchet differentiable at F, (3.1) holds, and*

$$E[\psi_F(X_1, X_2)]^2 < \infty \quad \textit{and} \quad E|\psi_F(X_1, X_1)| < \infty.$$

If $m = m_n \to \infty$ *and* $m_n/n \to 0$, *then* $\widehat{H}_{n,m}$ *is weakly consistent, i.e.,* (2.8) *holds.*

Proof. Define

$$V_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\psi_F(X_i, X_j), \qquad V_m^* = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\psi_F(X_i^*, X_j^*),$$

$$\widetilde{H}_n(x) = P\{nV_n \le x\}, \quad \text{and} \quad \widetilde{H}_{n,m}(x) = P_*\{m(V_m^* - V_n) \le x\}.$$

From the second-order differentiability of T,

$$n[T(F_n) - T(F)] = nV_n + o_p(1)$$

and, for any $\varepsilon > 0$,

$$P_*\{|m[T(F_m^*) - T(F_n)] - m(V_m^* - V_n)| > \varepsilon\} \to_p 0.$$

Hence, the result follows from

$$\sup_x |\widetilde{H}_{n,m}(x) - \widetilde{H}_n(x)| \to_p 0.$$

Let

$$\tilde{\psi}_n(x) = 2\left[\frac{1}{n}\sum_{j=1}^{n}\psi_F(x, X_j) - V_n\right].$$

Note that $\tilde{\psi}_n$ is the influence function of the "V-statistic" V_m^* and $\tilde{\psi}_n$ is not identically equal to zero, although the influence function of V_n is identically equal to zero. Let

$$\widetilde{V}_m^* = V_m^* - L_m^*,$$

where

$$L_m^* = \int \tilde{\psi}_n(x)\, d(F_m^* - F_n)(x) = \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\psi_F(X_i^*, X_j) - 2V_n.$$

Then $\widetilde{V}_m^*$ is a "V-statistic" with zero influence function. Using the same arguments in the proofs of Theorems 5.5.2 and 6.4.1B in [15], we can show that

$$\sup_x |P_*\{m(\widetilde{V}_m^* - V_m) \le x\} - \widetilde{H}_n(x)| \to_p 0.$$

Note that $m(V_m^* - V_m) = mL_m^* + m(\widetilde{V}_m^* - V_n)$. Hence the result follows if

$$P_*\{|mL_m^*| > \varepsilon\} \to_p 0 \tag{3.5}$$

for any $\varepsilon > 0$. Let E_* and var_* be the conditional expectation and variance taken under P_*. Since $E_*L_m^* = 0$, (3.5) follows from

$$\mathrm{var}_*(mL_m^*) \to_p 0. \tag{3.6}$$

Note that

$$\begin{aligned}\mathrm{var}_*(mL_m^*) &= \frac{4m}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}\psi_F(X_i, X_j) - V_n\right]^2 \\ &\le \frac{4m}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}\psi_F(X_i, X_j)\right]^2 \le 8(A_n + B_n),\end{aligned} \tag{3.7}$$

where

$$A_n = \frac{m^3}{n^3} \sum_{i=1}^{n} \left[\sum_{j \neq i} \psi_F(X_i, X_j) \right]^2 \quad \text{and} \quad B_n = \frac{m}{n^3} \sum_{i=1}^{n} [\psi_F(X_i, X_i)]^2 .$$

Note that

$$\begin{aligned} EA_n &= \frac{m}{n^2} E \left[\sum_{j=2}^{n} \psi_F(X_1, X_j) \right]^2 \\ &= \frac{m}{n^2} E \left\{ E \left[\sum_{j=2}^{n} \psi_F(x, X_j) \right]^2 \middle| X_1 = x \right\} \\ &= \frac{m(n-1)}{n^2} E \left\{ \int [\psi_F(X_1, y)]^2 \, dF(y) \right\}, \end{aligned}$$

where the last equality follows from the fact that $\int \psi_F(x, y) \, dF(y) = 0$ for all x. Thus, by $m/n \to 0$ and $E[\psi_F(X_1, X_2)]^2 < \infty$, we have $A_n \to_p 0$. Since $m/n \to 0$ and $E|\psi_F(X_1, X_1)| < \infty$, we also have $B_n \to_p 0$. Then (3.6) follows from $A_n \to_p 0$, $B_n \to_p 0$, and (3.7). □

4. Nondifferentiable functions of means

Consider nonsmooth $T_n = g(\overline{X}_n)$ with nondifferentiable g. For simplicity we focus on the case of $k = 1$. Assume that

$$\lim_{t \to 0\pm} \frac{g(\mu + t) - g(\mu)}{t} = g'(\mu\pm)$$

exist but $g'(\mu+) \neq g'(\mu-)$. An example is $g(x) = |x|$, which is nondifferentiable at 0 and $g'(0\pm) = \pm 1$.

By the mean-value theorem,

$$\begin{aligned} \sqrt{n}(T_n - \theta) = {} & g'(\mu+)\sqrt{n}(\overline{X}_n - \mu) I\{\overline{X}_n \geq \mu\} \\ & + g'(\mu-)\sqrt{n}(\overline{X}_n - \mu) I\{\overline{X}_n < \mu\} + o_p(n^{-1/2}). \end{aligned}$$

In general, $\sqrt{n}(T_n - \theta)$ is not asymptotically normal. In fact, the distribution

$$H_n(x) = P\left\{\sqrt{n}(T_n - \theta) \leq x\right\}$$

may not have any limit when both $g'(\mu+)$ and $g'(\mu-)$ are nonzero.

To see the inconsistency of the bootstrap estimator $\widehat{H}_n(x) = P_*\{\sqrt{n}(T_n^* - T_n) \leq x\}$ based on $X_1^*, \ldots, X_n^*$, we consider the special case of $g(x) = |x|$ and $\mu = 0$. Note that $\sqrt{n}|\overline{X}_n| \to_d |Z_{\sigma^2}|$, where $\sigma^2 = \operatorname{var}(X_1)$, and $\sqrt{n}(\overline{X}_n^* - \overline{X}_n) \to_d Z_{\sigma^2}$ a.s. But conditional on $X_1, X_2, \ldots$, $\sqrt{n}\,\overline{X}_n^*$ has no limit, since $\sqrt{n}\,\overline{X}_n$ does not converge. Thus,

$$\sqrt{n}(|\overline{X}_n^*| - |\overline{X}_n|) = \begin{cases} \sqrt{n}(\overline{X}_n^* - \overline{X}_n) - 2\sqrt{n}\,\overline{X}_n^* I\{\overline{X}_n^* < 0\}, & \overline{X}_n \geq 0, \\ \sqrt{n}(\overline{X}_n - \overline{X}_n^*) + 2\sqrt{n}\,\overline{X}_n^* I\{\overline{X}_n^* \geq 0\}, & \overline{X}_n < 0, \end{cases}$$

has no limit.

The reason for the inconsistency of the bootstrap estimators in this case is that $\overline{X}_n$ is not exactly equal to μ and both $\overline{X}_n$ and its bootstrap analogue $\overline{X}_n^*$ oscillate around the discontinuity point of g' with the same rate $n^{-1/2}$. The inconsistency can be rectified if the bootstrap analogue oscillates with a rate slower than $n^{-1/2}$. This again leads to the consideration of taking a bootstrap sample $X_1^*, \dots, X_m^*$ with $m/n \to 0$. Define $T_m^* = g(\overline{X}_m^*)$ and

$$\widehat{H}_{n,m}(x) = P_*\left\{\sqrt{m}(T_m^* - T_n) \le x\right\}.$$

Theorem 3. *Suppose that* $\mathrm{var}(X_1) < \infty$ *and that* g *is differentiable except at* μ *and* $g'(\mu+) \ne g'(\mu-)$.

(i) *If* $m = m_n \to \infty$ *and* $m_n/n \to 0$, *then* $\widehat{H}_{n,m}$ *is weakly consistent, i.e.,* (2.8) *holds.*

(ii) *If* $m_n \to \infty$ *and* $m_n \log\log n/n \to 0$, *then* $\widehat{H}_{n,m}$ *is strongly consistent, i.e.,* (2.9) *holds.*

Proof. Without loss of generality we assume $\mu = 0$. For m satisfying the condition in (i) or (ii),

$$\sqrt{m}(T_n - \theta) \to_p 0 \quad \text{or} \quad \to 0 \text{ a.s.}$$

Then

$$\begin{aligned}
\widehat{H}_{n,m}(x) &= P_*\{\sqrt{m}(T_m^* - \theta) \le x\} + o(1) \\
&= P^*\{g'(0+)\sqrt{m}\,\overline{X}_m^* \le x,\ \overline{X}_m^* \ge 0\} \\
&\quad + P_*\{g'(0-)\sqrt{m}\,\overline{X}_m^* \le x,\ \overline{X}_m^* < 0\} + o(1) \\
&= P\{g'(0+)\sqrt{n}\,\overline{X}_n \le x, \overline{X}_n \ge 0\} + P\{g'(0-)\sqrt{n}\,\overline{X}_n \le x,\ \overline{X}_n < 0\} + o(1) \\
&= P\{\sqrt{n}(T_n - \theta) \le x\} + o(1),
\end{aligned}$$

where the third equality follows from the fact that $\sqrt{m}(\overline{X}_m^* - \overline{X}_n)$ has the same limit as $\sqrt{n}\,\overline{X}_n$ [5] and the $o(1)$ is $o_p(1)$ for m satisfying the condition in (i) or $o(1)$ a.s. for m satisfying the condition in (ii). □

We have actually shown an example where H_n may not have any limit, but its bootstrap approximation $\widehat{H}_{n,m}$ is consistent.

5. Estimators based on tests

We now consider a type of nonsmooth estimators based on some test results. For illustration, we first consider the simple case where $X_1, \dots, X_n$ are i.i.d. random variables with mean μ and variance σ^2. Note that a large sample 2α-level test for the hypothesis $\mu = 0$ has rejection region

$$|\overline{X}_n| > z_\alpha S_n, \tag{5.1}$$

where $S_n^2 = n^{-2}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ and z_α is the $(1-\alpha)$th quantile of the standard normal distribution. Thus, an estimator of μ based on the test described by (5.1) is defined by

$$T_n = \begin{cases} \overline{X}_n, & |\overline{X}_n| > z_\alpha S_n, \\ 0, & |\overline{X}_n| \le z_\alpha S_n. \end{cases}$$

Let

$$H_n(x) = P\{\sqrt{n}(T_n - \mu) \le x\}.$$

When $\mu \neq 0$, $P\{|\overline{X}_n| \leq z_\alpha S_n\} \to 0$, and therefore

$$H_n(x) = P\{\sqrt{n}(\overline{X}_n - \mu) \leq x\} + o(1) \to \Phi(x/\sigma),$$

where $\Phi(x)$ is the standard normal distribution function. When $\mu = 0$,

$$P\{|\overline{X}_n| \leq z_\alpha S_n\} \to 1 - 2\alpha \tag{5.2}$$

and $H_n(x)$ converges weakly to a distribution $H(x)$ which is symmetric about zero and

$$H(x) = \begin{cases} 1 - 2\alpha, & 0 \leq x \leq z_\alpha \sigma, \\ \Phi(x/\sigma), & x > z_\alpha \sigma. \end{cases}$$

Let $X_1^*, \dots, X_m^*$ be i.i.d. bootstrap samples from F_n. If $m = n$, then the bootstrap analogue of T_n is

$$T_n^* = \begin{cases} \overline{X}_n^*, & |\overline{X}_n^*| > z_\alpha S_n^*, \\ 0, & |\overline{X}_n^*| \leq z_\alpha S_n^*, \end{cases}$$

where $S_n^{*2} = n^{-2} \sum_{i=1}^n (X_i^* - \overline{X}_n^*)^2$. The bootstrap estimator of H_n is then

$$\widehat{H}_n(x) = P_*\{\sqrt{n}(T_n^* - T_n) \leq x\}.$$

Suppose that $|\overline{X}_n| \leq z_\alpha S_n$. Then for $0 < x < z_\alpha \sigma$,

$$\begin{aligned} \widehat{H}_n(x) &= P_*\{\sqrt{n} T_n^* \leq x\} \\ &= P_*\{|\overline{X}_n^*| \leq z_\alpha S_n^*\} + P_*\{\sqrt{n}\,\overline{X}_n^* \leq x,\ |\overline{X}_n^*| > z_\alpha S_n^*\}. \end{aligned} \tag{5.3}$$

By Theorem 2.1 in [5] and $x < z_\alpha \sigma$,

$$P_*\{\sqrt{n}\,\overline{X}_n^* \leq x,\ |\overline{X}_n^*| > z_\alpha S_n^*\} \leq P_*\{z_\alpha \sqrt{n} S_n^* \leq x\} \to 0 \quad \text{a.s.} \tag{5.4}$$

and

$$\begin{aligned} P_*\{|\overline{X}_n^*| \leq z_\alpha S_n^*\} &= P_*\left\{-z_\alpha - \frac{\overline{X}_n}{S_n^*} \leq \frac{\overline{X}_n^* - \overline{X}_n}{S_n^*} \leq z_\alpha - \frac{\overline{X}_n}{S_n^*}\right\} \\ &= \Phi\left(z_\alpha - \frac{\sqrt{n}\,\overline{X}_n}{\sigma}\right) - \Phi\left(-z_\alpha - \frac{\sqrt{n}\,\overline{X}_n}{\sigma}\right) + o(1) \quad \text{a.s.} \end{aligned} \tag{5.5}$$

Since $\sqrt{n}\overline{X}_n$ is asymptotic nondegenerate when $\mu = 0$, the bootstrap estimator $\widehat{H}_n$ is inconsistent in view of (5.2)–(5.5).

The problem in this case is very similar to those in the previous sections: when $\mu = 0$, $\overline{X}_n$ does not equal 0 exactly, which leads to result (5.5), inconsistent with (5.2).

Now, we study the bootstrap with $m \neq n$ in a more general setting. Let $X_1, \dots, X_n$ be i.i.d. random k-vectors and $X_i = (X_{i1}, X_{i2})'$, where X_{i1} is a random k_1-vector and X_{i2} is a random k_2-vector, $k_1 + k_2 = k$. Let $\overline{X}_n = n^{-1} \sum_{i=1}^n X_i$, $S_n^2 = n^{-2} \sum_{i=1}^n (X_i - \overline{X}_n)(X_i - \overline{X}_n)'$, and $\overline{X}_{nj}$, S_{nj}^2, $j = 1, 2$, be similarly defined for subvectors X_{ij}, $i = 1, \dots, n$. Let $\mu_j = EX_{ij}$. A large sample 2α-level test for $\mu_1 = 0$ has rejection region

$$\overline{X}_{n1}'(S_{n1}^2)^{-1}\overline{X}_{n1} > \chi_\alpha^2(k_1),$$

where $\chi^2_\alpha(k_1)$ is the $(1-\alpha)$th quantile of the chi-square distribution with k_1 degrees of freedom. Consider the estimator of $\mu = EX_i$ based on the result of the test of $\mu_1 = 0$:

$$T_n = \begin{cases} \overline{X}_n, & \overline{X}'_{n1}(S^2_{n1})^{-1}\overline{X}_{n1} > \chi^2_\alpha(k_1), \\ (0, \overline{X}_{n2})', & \overline{X}'_{n1}(S^2_{n1})^{-1}\overline{X}_{n1} \le \chi^2_\alpha(k_1). \end{cases}$$

Let H_n be the k-dimensional distribution of $\sqrt{n}(T_n - \mu)$. Let $X^*_1, \dots, X^*_m$ be i.i.d. bootstrap samples from the empirical distribution F_n; $\overline{X}^*_m$, S^{*2}_m, $\overline{X}^*_{mj}$, S^{*2}_{mj}, and T^*_m be the bootstrap analogues of $\overline{X}_n$, S^2_n, $\overline{X}_{nj}$, S^2_{nj}, and T_n, respectively; and $\widehat{H}_{n,m}$ be the conditional (given $X_1, \dots, X_n$) distribution of $\sqrt{m}(T^*_m - T_n)$.

Theorem 4. *Suppose that* $\operatorname{var}(X_1) < \infty$.

(1) *If* $m = m_n \to \infty$ *and* $m_n/n \to 0$, *then* $\widehat{H}_{n,m}$ *is weakly consistent, i.e.,* (2.8) *holds.*

(ii) *If* $m_n \to \infty$ *and* $m_n \log\log n/n \to 0$, *then* $\widehat{H}_{n,m}$ *is strongly consistent, i.e.,* (2.9) *holds.*

Proof. Without loss of generality we may assume $k = k_1$, i.e., $X_i = X_{i1}$. Consider the case of $\mu \ne 0$. Then by a multivariate version of Theorem 2.1 in [5],

$$P_*\{(\overline{X}^*_m)'(S^{*2}_m)^{-1}\overline{X}^*_m \le \chi^2_\alpha(k)\} \to 0 \quad \text{a.s.},$$

and therefore the result follows.

Now assume $\mu = 0$. Then for m satisfying the condition in (i) or (ii),

$$\sqrt{m}\|T_n\| \le \sqrt{m}\|\overline{X}_n\| \to_p 0 \quad \text{or} \quad \to 0 \text{ a.s.}, \tag{5.6}$$

where $\|x\| = \sqrt{x'x}$. Let A_x denote the subset of $\mathbb{R}^k$ of the form

$$(-\infty, x_1] \times \cdots \times (-\infty, x_k],$$

where x_j is the jth component of $x \in \mathbb{R}^k$. In the following proof, $o(1)$ is either $o_p(1)$ or $o(1)$ a.s., depending on whether m satisfies the condition in (i) or (ii). Suppose that $0 \in A_x$. Then

$$\begin{aligned}
&P_*\{\sqrt{m}(T^*_m - T_n) \in A_x\} \\
&\quad = P_*\{\sqrt{m}T^*_m \in A_x, (\overline{X}^*_m)'(S^{*2}_m)^{-1}\overline{X}^*_m > \chi^2_\alpha(k)\} \\
&\qquad + P_*\{\sqrt{m}T^*_m \in A_x, (\overline{X}^*_m)'(S^{*2}_m)^{-1}\overline{X}^*_m \le \chi^2_\alpha(k)\} + o(1) \\
&\quad = P_*\{\sqrt{m}\overline{X}^*_m \in A_x, (\overline{X}^*_m)'(S^{*2}_m)^{-1}\overline{X}^*_m > \chi^2_\alpha(k)\} \\
&\qquad + P_*\{(\overline{X}^*_m)'(S^{*2}_m)^{-1}\overline{X}^*_m \le \chi^2_\alpha(k)\} + o(1) \\
&\quad = P_*\{\sqrt{m}(\overline{X}^*_m - \overline{X}_n) \in A_x, (\overline{X}^*_m - \overline{X}_n)'(S^{*2}_m)^{-1}(\overline{X}^*_m - \overline{X}_n) > \chi^2_\alpha(k)\} \\
&\qquad + P_*\{(\overline{X}^*_m - \overline{X}_n)'(S^{*2}_m)^{-1}(\overline{X}^*_m - \overline{X}_n) \le \chi^2_\alpha(k)\} + o(1) \\
&\quad = P\{\sqrt{n}\overline{X}_n \in A_x, \overline{X}'_n(S^2_n)^{-1}\overline{X}_n > \chi^2_\alpha(k)\} + 1 - 2\alpha + o(1) \\
&\quad = P\{\sqrt{n}T_n \in A_x\} + o(1),
\end{aligned}$$

where we have repeatedly used (5.6) and Theorem 2.1 in [5]. A similar result for the case of $0 \notin A_x$ can also be established. This proves the result. □

6. Discussions

There are other examples similar to those we studied in §§2–5.

Extreme order statistics. Bickel and Freedman [5] and Loh [14] showed that the bootstrap estimators of the distributions of the extreme-order statistics are inconsistent. The reason for the inconsistency of the bootstrap in this case is the same as that in §§2–5, although the situations are entirely different. Let $X_{(n)}$ be the maximum of i.i.d. random variables $X_1, \dots, X_n$ from F with $F(\theta) = 1$ for some θ, and let $X^*_{(m)}$ be the maximum of $X^*_1, \dots, X^*_m$ which are i.i.d. from the empirical distribution F_n. Although $X_{(n)} \to \theta$, it never equals θ. But

$$P_*\{X^*_{(n)} = X_{(n)}\} = 1 - (1 - n^{-1})^n \to 1 - e^{-1}, \tag{6.1}$$

which leads to the inconsistency of the bootstrap estimator. Bickel and Freedman [5] pointed out that this inconsistency cannot be mended by smoothing, i.e., taking bootstrap samples from a smooth estimator $\widetilde{F}_n$ of F. However, with a bootstrap sample size m satisfying $m/n \to 0$, $X_{(n)}$ converges to θ faster than its bootstrap analogue $X^*_{(m)}$ and, in contrast with (6.1),

$$P_*\{X^*_{(m)} = X_{(n)}\} = 1 - (1 - n^{-1})^m \to 0.$$

Indeed, Swanepoel [17] and Deheuvels, Mason, and Shorack [6] showed that the bootstrap estimator is weakly consistent if $m = m_n \to \infty$ and $m/n \to 0$ and is strongly consistent if $m \to \infty$ and $m \log\log n/n \to 0$.

Sample quantiles. Consider the sample quantile

$$T_n = F_n^{-1}(p) = \inf\{x : F_n(x) \geq p\},$$

where $0 < p < 1$ and F_n is the empirical distribution of i.i.d. random variables $X_1, \dots, X_n$. Let θ be the population pth quantile and H_n be the distribution of $R_n = \sqrt{n}(T_n - \theta)$. Under the regularity condition that the population distribution F is differentiable at θ and $f(\theta) > 0$, $f(x) = dF/dx$, the bootstrap estimator $\widehat{H}_n$ of H_n is strongly consistent (e.g., [5]). However, if

$$\lim_{t \to 0\pm} \frac{F(\theta + t) - F(\theta)}{t} = f(\theta\pm)$$

exist and are positive but $f(\theta-) \neq f(\theta+)$, then H_n still has a limit (e.g., [15, p. 77]); the bootstrap estimator with $m = n$ is inconsistent; and the bootstrap estimator is weakly consistent if $m = m_n \to \infty$ and $m/n \to 0$ and is strongly consistent if $m \to \infty$ and $m \log\log n/n \to 0$ [13, Theorem 2]. Similar results in the case where $f(\theta-) = f(\theta+) = f(\theta)$ but $f(\theta) = 0$ can also be obtained (see [13]).

Variable selection. Another example is given in [16] in the context of variable selection in regression problems. It was found that when the bootstrap sample size is n, the variable selection procedure based on minimizing the bootstrap estimator of expected excess error [8, p. 52] is inconsistent, whereas using the bootstrap sample size m satisfying $m \to \infty$ and $m/n \to 0$, we can construct a consistent bootstrap variable selection procedure.

REFERENCES

1. L. Abramovitch and K. Singh, *Edgeworth corrected pivotal statistics and the bootstrap*, Ann. Statist. **13** (1981), 116–132.
2. K. B. Athreya, *Bootstrap of the mean in the infinite variance case*, Ann. Statist. **15** (1987), 724–731.
3. G. J. Babu, *Bootstrapping statistics with linear combinations of chi-squares as weak limit*, Sankhyā Ser. A **46** (1984), 85–93.
4. R. Beran, *Prepivoting to reduce level error of confidence sets*, Biometrika **74** (1987), 457–468.
5. P. J. Bickel and D. A. Freedman, *Some asymptotic theory for the bootstrap*, Ann. Statist. **9** (1981), 1196–1217.
6. P. Deheuvels, D. M. Mason, and G. R. Shorack, *Some results on the influence of extremes on the bootstrap*, Ann. Inst. Henri Poincaré **29** (1993), 83–103.
7. B. Efron, *Bootstrap methods: another look at the jackknife*, Ann. Statist. **7** (1979), 1–26.
8. ——, *The jackknife, the bootstrap, and other resampling plans*, SIAM, Philadelphia, 1982.
9. ——, *Better bootstrap confidence intervals* (*with discussion*), J. Amer. Statist. Assoc. **82** (1987), 171–200.
10. P. Hall, *Theoretical comparison of bootstrap confidence intervals* (*with discussion*), Ann. Statist. **16** (1988), 927–985.
11. ——, *On the relative performance of bootstrap and Edgeworth approximations of a distribution function*, J. Multivariate Anal. **35** (1990), 108–129.
12. F. R. Hampel, *The influence curve and its role in robust estimation*, J. Amer. Statist. Assoc. **69** (1974), 383–393.
13. J. S. Huang, P. K. Sen, and J. Shao, *Sample quantiles in some nonregular cases and the vulnerable bootstrap*, preprint, 1992.
14. W.-Y. Loh, *Estimating an endpoint of a distribution with resampling methods*, Ann. Statist. **12** (1984), 1543–1550.
15. R. J. Serfling, *Approximation theorems of mathematical statistics*, Wiley, New York, 1980.
16. J. Shao, *Bootstrap variable selection in regression*, preprint, 1992.
17. J. W. H. Swanepoel, *A note on proving that the* (*modified*) *bootstrap works*, Comm. Statist. A—Theory Methods **15** (1986), 3193–3203.

DEPARTMENT OF STATISTICS, UNIVERSITY OF WISCONSIN, MADISON, WISCONSIN 53706
E-mail address: `shao@atlas.stat.wisc.edu`