The Annals of Statistics
1984, Vol. 12, No. 3, 1130–1135

A NOTE ON BOOTSTRAPPING THE SAMPLE MEDIAN

By Malay Ghosh[1], William C. Parr, Kesar Singh[2] and G. Jogesh Babu

University of Florida; University of Florida; Rutgers University; and Indian Statistical Institute, Calcutta

Efron (1979, 1982), in his treatment of the bootstrap, discusses its use for estimation of the asymptotic variance of the sample median, in the sampling situation of independent and identically distributed random variables with common distribution function F having a positive derivative continuous in a neighborhood of the true median μ. The natural conjecture that the bootstrap variance estimator converges almost surely to the asymptotic variance is shown by an example to be false unless a tail condition is imposed on F. We prove that such strong convergence does hold under the fairly nonrestrictive condition that $E[|X^\alpha] < \infty$ for some $\alpha > 0$.

1. Introduction and notation. Throughout this paper we consider observing $X_1, X_2, \cdots, X_n$, a random sample from a univariate distribution with distribution function F having a positive derivative f continuous in a neighborhood of its median $\mu = \inf\{t \mid F(t) \geq ½\}$. Let $F_n(t) = \sum_{i=1}^n I(X_i \leq t)/n$ for all real t be the ordinary empirical distribution function. Define the sample median as $m_n = \inf\{t \mid F_n(t) \geq ½\}$. Under the conditions stated above (and even slight weakenings thereof) we have that

$$\sqrt{n}(m_n - \mu) \to N(0, \sigma^2), \tag{1.1}$$

where $\sigma^2 = 1/(4f^2(\mu))$.

Two methods in common use for the nonparametric estimation of standard errors are the jackknife (see Miller, 1974) and the bootstrap (see Efron, 1979, 1982). Even under the smoothness conditions stated above, it can be shown that the jackknife estimator of σ^2 has the undesirable property of converging in law (along a sequence of even sample sizes) to a *random* variable which has the distribution of $(1/4f^2(\mu))(W/2)^2$, where W has a chi-squared distribution with 2 degrees of freedom. The fact that the jackknife fails so dramatically in this situation has been viewed as a kind of "smoking gun" for the bootstrap estimator of variance, which has been presumed to perform satisfactorily in this problem.

The bootstrap estimator of the asymptotic variance of $Z_n = \sqrt{n}(m_n - \mu)$ may be motivated as follows. (See Efron, 1982, page 27, for a more detailed explanation.) Compute theoretically the variance of $Z_n^* = \sqrt{n}(m_n^* - m_n)$, where m_n^* is the median of a random sample of size n drawn with replacement from the original sample $X_1, \cdots, X_n$ and the variance is computed conditional on the

Received September 1983; revised February 1984.

[1] Research supported by the Army Research Office, Durham Grant Number DAAG29-82-K-0136.

[2] Research supported by NSF Grant MCS 83-01082.

AMS 1980 *subject classifications.* Primary 62E20; secondary 62G05.

Key words and phrases. Bootstrap, median, standard error estimation.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

observed values of $X_1, \cdots, X_n$. A simple calculation shows this to be

$$\hat{\sigma}_n^2 = n\{\sum_{i=1}^n p_n(i) Y_i^2 - (\sum_{j=1}^n p_n(j) Y_j)^2\}.$$
$$= \sum_{i=1}^n p_n(i)\{\sqrt{n}(Y_i - m_n)\}^2 - (\sum_{j=1}^n p_n(j)\sqrt{n}(Y_j - m_n))^2,$$

where Y_i is the ith order statistic of $X_1, \cdots, X_n$ (we omit the second subscript indexing the sample size for notational compactness), and

$$p_n(i) = \sum_{j=0}^{a-1}\left[b_{j,n}\left(\frac{i-1}{n}\right) - b_{j,n}\left(\frac{i}{n}\right)\right], \quad \text{for} \quad i \geq 2$$
$$= 1 - \sum_{j=0}^{a-1} b_{j,n}\left(\frac{1}{n}\right) \qquad \text{for} \quad i = 1$$

where $b_{j,n}(u) = \binom{n}{j}u^j(1-u)^{n-j}$, and $a = [n/2] + 1$.

Bickel and Freedman (1981) (see also Theorem 2 of Singh, 1981) have shown the following:

LEMMA 1. (Their Proposition 5.1). *Suppose F has a unique median μ and F has a derivative f positive and continuous in a neighborhood of μ. Then along almost all sample sequences $X_1, X_2, \cdots$, the conditional law of $\sqrt{n}(m_n^* - m_n)$, given $X_1, \cdots, X_n$ converges weakly to $N(0, \sigma^2)$, with $\sigma^2 = 1/(4f^2(\mu))$ as above.*

It is thus natural to wonder whether $\hat{\sigma}_n^2 \to \sigma^2$ almost surely, or at least in probability. Such is not the case, however, without further conditions.

EXAMPLE. Let $F(x)$ be such that $F(x) = 1 - F(-x)$, with derivative $F' = f$ continuous and positive in a neighborhood of 0, but with $F(x) = 1 - (\ell_2(x))^{-1}$, for $x > C$, for some large positive C, where $\ell_2(x) = \log\log x$ if $x > e$, and $\ell_2(x) = 1$, otherwise. Then, $\hat{\sigma}_n^2 \to \infty$ almost surely.

The following lemma is helpful.

LEMMA 2. *Let $U_1, U_2, \cdots$ be such that*

(a) *$\{U_n\}_{n=1}^\infty$ is tight, and*

(b) *$E[U_n^2] \to \infty$, but $E[U_n^2] < \infty$ for all n.*

Then

$$\operatorname{Var}(U_n) \to \infty.$$

PROOF. Assume there exists an infinite subsequence $n_1 < n_2 < n_3 < \cdots$ such that $\operatorname{Var}(U_{n_i}) < B < \infty$ for all i. But then, since $\operatorname{Var}(U_{n_i}) = E[U_{n_i}^2] - (E[U_{n_i}])^2$, and $E[U_{n_i}^2] \to \infty$, we must have that $|E[U_{n_i}]| \to \infty$. Also, for any $K > 1$,

$$P[|U_{n_i} - E[U_{n_i}]| > K\sqrt{B}] < 1/K^2,$$

since $B \geq \operatorname{Var}(U_{n_i})$. Since $E(U_{n_i}) \to \infty$, pick j such that $|E[U_{n_j}]| > 3K\sqrt{B}$ for all $i \geq j$. Therefore,

$$P\{U_{n_i} \in [-2K\sqrt{B}, 2K\sqrt{B}]\} < 1/K^2$$

for all $i \geq j$. But K was arbitrarily large, and the above thus contradicts tightness of the $\{U_n\}$. Therefore, $\operatorname{Var}(U_n) \to \infty$.

Using this lemma, we thus need only show that $E_*[n(m_n^* - m_n)^2] \to \infty$ almost surely (a.s.), where E_* refers to expectation over the conditional law of m_n^*, given $X_1, X_2 \cdots X_n$. Let $Y_1, Y_2 \cdots Y_n$ denote the order statistics of $X_1, X_2 \cdots X_n$. Since Y_n can appear at each of the second stage draws with a chance $(1/n)^n$, it follows that

$$E_*[n(m_n^* - m_n)^2] \geq n^{-n+1}(Y_n - m_n)^2.$$

The population median has been assumed to be zero; therefore $m_n \to 0$ a.s. Thus, the claim $\hat{\sigma}_n^2 \to \infty$ a.s. follows if we show that $n^{-n+1}Y_n^2 \to \infty$ a.s. For any constant $K > 0$,

$$\begin{aligned} P(n^{-n+1}Y_n^2 < K) &\leq P(Y_n < K^{1/2}n^{(n-1)/2}) \\ &= [1 - [\ell_2(K^{1/2}n^{(n-1)/2})]^{-1}]^n \\ &\leq \exp\{-n[\log((n-1)/2)\log n + (1/2)\log K]^{-1}\} \\ &\leq \exp\{-n/K_1 \log n\} \end{aligned}$$

for large n and some $K_1 > 0$. Thus, $n^{-n+1}Y_n^2 \to \infty$ a.s. in view of the Borel Cantelli lemma.

It is clear from the above example that at least some tail condition is needed on F to ensure consistency of the bootstrap estimator of variance. The following section shows that $\hat{\sigma}_n^2 = V_*[\sqrt{n}(m_n^* - m_n)]$ converges almost surely to $(4f^2(\mu))^{-1}$ under a very nonrestrictive moment assumption on the X_i's. In the above, V_* denotes variance over the conditional law of m_n^* given $X_1, \cdots, X_n$.

2. The main result. First, we state a lemma needed in proving the main result. The lemma is known in the literature. A proof is included here for the sake of completeness.

LEMMA 3. *Let $X_1, \cdots, X_n$ be* iid *such that $E|X_1|^\alpha < \infty$ for some $\alpha > 0$. Let $Y_1 \leq \cdots \leq Y_n$ denote the ordered X_i's. Then, $(|Y_n| + |Y_1|)/n^{1/\alpha} \to 0$ a.s.*

PROOF. $E|X_1|^\alpha < \infty$ implies that, for every $\varepsilon > 0$,

$$\textstyle\sum_1^\infty P(|X_i| > \varepsilon i^{1/\alpha}) = \sum_1^\infty P(|X_1| > \varepsilon i^{1/\alpha}) < \infty.$$

So, in view of the Borel-Cantelli lemma, $|X_i| < \varepsilon i^{1/\alpha}$ for all but finitely many i's, a.s. Hence, $(|Y_1| + |Y_n|)/n^{1/\alpha} \to 0$ a.s.

We are now in a position to prove the main result, namely the strong consistency of the bootstrap variance estimator under an extremely weak moment condition.

THEOREM 1. *Let* $X_1, \cdots, X_n$ be iid *with* $E|X_1|^{\alpha} < \infty$ *for some* $\alpha > 0$. *Also, let the conditions of Lemma* 1 *hold. Then,* $\hat{\sigma}_n^2 = V_*[\sqrt{n}(m_n^* - m_n)] \to_{\text{a.s.}} (4f^2(\mu))^{-1}$ *as* $n \to \infty$, *that is, the bootstrap variance estimator is strongly consistent.*

PROOF. In view of Lemma 1, it suffices to prove the uniform integrability of $n(m_n^* - m_n)^2$ (which implies uniform integrability of $\sqrt{n}(m_n^* - m_n)$ as well). It suffices for this to show that $E_* | \sqrt{n}(m_n^* - m_n)|^{2+\delta} < \infty$ for some $\delta > 0$. Next, denoting by P^* the conditional probability law of m_n^* given $X_1, \cdots, X_n$, it follows that

$$E_* | \sqrt{n}(m_n^* - m_n)|^{2+\delta} = (1 + \delta) \int_0^{\infty} t^{1+\delta} P_*(\sqrt{n}\,| m_n^* - m_n | > t]\, dt.$$

Thus, it suffices to show that, for a constant $c > 0$, all $t > 1$ and a $\delta' > 0$,

$$P_*(\sqrt{n}\,| m_n^* - m_n | > t) \le ct^{-(2+\delta')} \tag{2.1}$$

for all large n, a.s.

In order to establish this, we argue separately in two different zones (I) $t \in [1, c(\alpha)(\log n)^{1/2}]$ and (II) $[c(\alpha)(\log n)^{1/2}, \infty)$ where the requirement on the constant $c(\alpha)$ is specified later.

$$\begin{aligned} &\{\sqrt{n}(m_n^* - m_n) > t\} \\ &\equiv \{\tfrac{1}{2} + 1/2n \ge F_n^*(m_n + t/\sqrt{n})\} \qquad (2.2)\\ &\equiv \{\tfrac{1}{2} + 1/2n - F_n(m_n + t/\sqrt{n}) \ge F_n^*(m_n + t/\sqrt{n}) - F_n(m_n + t/\sqrt{n})\} \end{aligned}$$

where F_n^* denotes the bootstrap empirical c.d.f. and F_n is the usual empirical c.d.f. based on the X_i's. Let us write

$$\begin{aligned} \tfrac{1}{2} + 1/(2n) - F_n(m_n + t/\sqrt{n}) &= [F(m_n) - F(m_n + t/\sqrt{n})] \\ &\quad + [F(m_n + t/\sqrt{n}) - F_n(m_n + t/\sqrt{n}) - F(m_n) + F_n(m_n)] \\ &\quad + [\tfrac{1}{2} + 1/(2n) - F_n(m_n)] \\ &= A_n + B_n + C_n \quad \text{(say)}. \end{aligned}$$

Because of the assumed continuity of F in a neighborhood of μ, it follows that $|C_n| \le 1/n$ for all large n, a.s. Using Lemma 1 of Bahadur (1966) and the well-known fact that $|m_n - \mu| = O(n^{-1/2}(\log n)^{1/2})$ a.s., we deduce that $|B_n| = O(n^{-3/4}\log n)$ a.s. in the region $t \le c(\alpha)(\log n)^{1/2}$. Also, in this region of t, it is clear from Taylor's expansion that $A_n = -(t/\sqrt{n})F'(m_n) + o(n^{-1/2}(\log n)^{1/2})$ a.s. Combining all the above facts, and noting that $t > 1$, we conclude that, for all large n,

$$\tfrac{1}{2} - F_n(m_n + t/\sqrt{n}) \le -\varepsilon t/\sqrt{n} \tag{2.3}$$

for some $\varepsilon > 0$ and all $t \in (1, c(\alpha)(\log n)^{1/2}]$, a.s. Consequently, it follows from

Markov's inequality that, in zone (I) of t,

$$P_*(\sqrt{n}(m_n^* - m_n) > t) \le (\varepsilon t)^{-4} E_*[\sqrt{n}(F_n^*(m_n + t/\sqrt{n}) - F_n(m_n + t/\sqrt{n}))]^4$$
$$\le 3(\varepsilon t)^{-4}.$$

For $t > c(\alpha)(\log n)^{1/2}$, using (2.2) and (2.3), for all large n a.s.,

$$\begin{aligned} &P_*(\sqrt{n}(m_n^* - m_n) > t) \\ &\quad \le P_*(\sqrt{n}(m_n^* - m_n) > c(\alpha)(\log n)^{1/2}) \\ &\quad \le P_*(F_n^*(m_n + c(\alpha)(\log n)^{1/2} n^{-1/2}) - F_n(m_n + c(\alpha)(\log n)^{1/2} n^{-1/2}) \\ &\qquad \le -\varepsilon c(\alpha)(\log n)^{1/2} n^{-1/2}). \end{aligned} \tag{2.4}$$

Choose $c(\alpha) = 1/\alpha + ½$. Now it follows from Lemma 3.1 of Singh (1981) with $p = F_n(m_n + c(\alpha)(\log n)^{1/2} n^{-1/2})$, $B = 1$, $Z = (1/\alpha + ½)(2 + \delta)\log n$, $D = \varepsilon c(\alpha)(1 + e/2)^{-1}(\log n)^{1/2} n^{1/2}$ that the right-hand side of (2.4) is $O(n^{-(1/\alpha + 1/2)(2+\delta)})$. Hence, for large n,

$$\int_{c(\alpha)(\log n)^{1/2}}^{n^{1/\alpha+1/2}} t^{1+\delta} P_*(\sqrt{n}(m_n^* - m_n) > t)\, dt = O(1) \quad \text{a.s.}$$

Finally, because of Lemma 3, $P_*(\sqrt{n}(m_n^* - m_n) > n^{1/2+1/\alpha}) = 0$ a.s. for all large n. Thus, we have proved (2.1) with $\sqrt{n}\,|m_n^* - m_n|$ replaced by $\sqrt{n}(m_n^* - m_n)$. Similar arguments can be used to handle $-\sqrt{n}(m_n^* - m_n)$. This completes the proof of (2.1).

3. Some remarks. An examination of the proof of Theorem 1 suggests the following "robustification" of the bootstrap to avoid the (admittedly nononerous) moment condition. First, Winsorize the original sample, replacing Y_i by $Y_{[nz]}$ for all $i \le [nz]$, and by $Y_{[n(1-z)]}$ for all $i \ge [n(1 - z)]$, for some $z \in (0, ½)$. Then, perform the bootstrap on the modified sample. Examining the previous proof, we find that $\hat{\sigma}_n^2 \to_{\text{a.s.}} [4f^2(\mu)]^{-1}$, even without the moment condition, since the only use of the moment condition was to bound $|Y_1| + |Y_n|$ by n^β for some $\beta > 0$ and this term is now replaced by $|Y_{[nz]}| + |Y_{[n(1-z)]}|$ which is trivially $O(1)$ a.s. An alternative to Winsorizing the original sample is trimming it. Draw the second stage samples from $Y_{[nz]}, Y_{[nz]+1}, \cdots Y_{[n(1-z)]}$ assigning each one equal probability $(= 1/[[n(1 - z)] - [nz] + 1])$ at each draw and define the estimator of the variance as follows:

$$\sigma_n^{*2} = \frac{1}{(1 - 2z)^2} E_*(m_n^* - m_n)^2.$$

Following the proof of Theorem 1, it is not hard to show that this σ_n^{*2} converges to $1/(4f^2(\mu))$ a.s., without requiring any moment condition.

Needless to say, all the above discussions extend appropriately to any general quantile without any further complication. In fact, with the help of the Kiefer type representation of quantile processes, one can extend Theorem 1 and the remarks of the previous paragraph to a general trimmed type L-statistic.

Acknowledgement. Thanks are due to Professor K. E. Athreya in connection with the proof of Lemma 2. Thanks are also due to Professor P. K. Sen for some useful comments.

REFERENCES

BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. (1982). *The Jackknife, The Bootstrap, and Other Resampling Schemes.* SIAM, Philadelphia.

MILLER, R. G. (1974). The jackknife—a review. *Biometrika* **61** 1–15.

SINGH, K. (1981). On asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

MALAY GHOSH
DEPARTMENT OF STATISTICS
NUCLEAR SCIENCES CENTER
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611

WILLIAM C. PARR
DEPARTMENT OF STATISTICS
NUCLEAR SCIENCES CENTER
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611

KESAR SINGH
DEPARTMENT OF STATISTICS
HILL CENTER
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903

G. JOGESH BABU
STAT. MATH. DIVISION
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD
CALCUTTA-700035, INDIA