Ann. Inst. Statist. Math.
Vol. 42, No. 4, 737–752 (1990)

BOOTSTRAP ESTIMATION OF THE ASYMPTOTIC VARIANCES OF STATISTICAL FUNCTIONALS

JUN SHAO

Department of Mathematics, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5

(Received January 11, 1989; revised August 25, 1989)

Abstract. A modified bootstrap estimator of the asymptotic variance of a statistical functional is studied. The modified bootstrap variance estimator circumvents the problem of the original bootstrap when the population distribution has heavy tails, and requires less stringent conditions for its consistency than the ordinary bootstrap variance estimator. The consistency of the modified bootstrap variance estimator is established for differentiable statistical functionals.

Key words and phrases: Bootstrap, variance estimation, Fréchet differentiability, consistency.

1. Introduction

Let θ be a parameter of interest and $\hat{\theta}_n$ be its estimator based on n observations. In many situations, $n^{1/2}(\hat{\theta}_n - \theta)$ has a limit distribution with a finite variance σ^2. A consistent estimator of σ^2 is needed for various purposes in statistical inference with a large sample size n. The bootstrap (Efron (1979)) is a widely applicable and convenient method of estimating σ^2. A detailed description of the bootstrap procedure is given in the next section. Throughout the paper we use $\hat{\sigma}_b^2$ to denote the bootstrap estimator of σ^2.

The bootstrap distribution, which approximates the distribution of $n^{1/2}(\hat{\theta}_n - \theta)$ (see Section 2), converges to the same limit distribution as $n^{1/2}(\hat{\theta}_n - \theta)$ under reasonable conditions (see Section 3). But this need not entail the consistency of $\hat{\sigma}_b^2$, which is the variance of the bootstrap distribution, since the variance functional is not weakly continuous. For some special types of estimators $\hat{\theta}_n$, such as the sample mean and sample median, the consistency of $\hat{\sigma}_b^2$ has been proved by Bickel and Freedman (1981), Singh (1981) and Ghosh *et al.* (1984) under some conditions. However, a general theory for the consistency of the bootstrap variance estimator is not available.

A modified bootstrap estimator $\hat{\sigma}_a^2$ is described in Section 2. The main purpose of this paper is to study the asymptotic properties of this modified bootstrap variance estimator for the case where $\theta = T(F)$ is a functional defined on a class of distributions F and $\hat{\theta}_n$ is T evaluated at F_n, the empirical distribution function of the observations. Precise definitions are given in Section 2.

We show in Section 2 that the modified bootstrap estimator $\hat{\sigma}_a^2$ has better asymptotic properties than the original bootstrap variance estimator $\hat{\sigma}_b^2$ (Theorem 2.1 and Example 2.1) and is asymptotically equivalent to $\hat{\sigma}_b^2$ in some cases (Example 2.2). In Section 3, $\hat{\sigma}_a^2$ is proved to be consistent if T is Fréchet differentiable (Theorem 3.1). By assuming that T has a stronger version of this differential, we prove the almost sure consistency of $\hat{\sigma}_a^2$ (Theorem 3.2).

2. The bootstrap and its modification

Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) real-valued random variables with an unknown distribution $F \in \Xi$, where Ξ is a convex class of distribution functions containing all degenerate distributions. Often the parameter of interest is $\theta = T(F)$, where T is a functional from Ξ to the real line, and the estimator of θ is $\hat{\theta}_n = T(F_n)$, where F_n is the empirical distribution function based on $X_1, \ldots, X_n$. Examples can be found in Serfling (1980). We assume that $n^{1/2}(\hat{\theta}_n - \theta)$ has a limit distribution with variance σ^2 (this is ensured by the differentiability of T (see Section 3)).

Let $X_1^*, \ldots, X_m^*$ be i.i.d. samples from the empirical distribution F_n, F_{nm}^* be the empirical distribution function based on $X_1^*, \ldots, X_m^*$, and $\hat{\theta}_{nm}^* = T(F_{nm}^*)$. Throughout the paper P_*, E_* and Var_* denote the bootstrap probability, expectation and variance (conditional on $X_1, \ldots, X_n$), respectively. The bootstrap approximation to the distribution of $n^{1/2}(\hat{\theta}_n - \theta)$ is the conditional distribution of $m^{1/2}(\hat{\theta}_{nm}^* - \hat{\theta}_n)$, and the bootstrap estimator of σ^2 is $\hat{\sigma}_b^2 = m \,\mathrm{Var}_* \,\hat{\theta}_{nm}^*$. Note that we allow the bootstrap sample size m to be different from n. For the advantages of having $m \neq n$, see Swanepoel (1986) and Rao and Wu (1988). To evaluate $\hat{\sigma}_b^2$, one may use the Monte Carlo approximation

$$B^{-1} \sum_{b=1}^{B} \left(\hat{\theta}_{nm}^{*b} - B^{-1} \sum_{b=1}^{B} \hat{\theta}_{nm}^{*b} \right)^2, \tag{2.1}$$

where $\hat{\theta}_{nm}^{*b} = T(F_{nm}^{*b})$ and F_{nm}^{*b} is the empirical distribution based on an i.i.d. sample $X_1^{*b}, \ldots, X_m^{*b}$ from F_n, $b = 1, \ldots, B$.

Even for differentiable T, $\hat{\sigma}_b^2$ may not be consistent. Counter-examples can be found in Ghosh *et al.* (1984) and Shao (1988*a*). We study the following modified bootstrap estimator:

(2.2) $$\hat{\sigma}_a^2 = m \operatorname{Var}_* [\Delta_{nm}^*(a_m)] ,$$

where $a_m \in (0, \infty]$ and $\Delta_{nm}^*(a_m)$ is obtained by truncating $\hat{\theta}_{nm}^* - \hat{\theta}_n$ at a_m and $-a_m$, i.e.,

(2.3) $$\Delta_{nm}^*(a_m) = \begin{cases} a_m & \text{if } \hat{\theta}_{nm}^* - \hat{\theta}_n > a_m \\ \hat{\theta}_{nm}^* - \hat{\theta}_n & \text{if } |\hat{\theta}_{nm}^* - \hat{\theta}_n| \le a_m \\ -a_m & \text{if } \hat{\theta}_{nm}^* - \hat{\theta}_n < -a_m . \end{cases}$$

In the special case of $a_m \equiv \infty$, $\hat{\sigma}_a^2$ reduces to $\hat{\sigma}_b^2$. The Monte Carlo approximation to $\hat{\sigma}_a^2$ can be obtained by using (2.1) and truncating $\hat{\theta}_{nm}^{*b} - \hat{\theta}_n$ at a_m and $-a_m$. The choice of a_m is discussed in Section 4.

Consistency of $\hat{\sigma}_a^2$ in the special cases where $\hat{\theta}_n$ is a differentiable function of the sample mean or sample quantile is proved in Shao (1988*a*).

In Theorem 2.1 we show that the consistency of $\hat{\sigma}_b^2$ implies the consistency of $\hat{\sigma}_a^2$. Example 2.1 indicates that $\hat{\sigma}_a^2$ improves $\hat{\sigma}_b^2$ when the distribution F has heavy tails. Hence $\hat{\sigma}_a^2$ has better asymptotic performance than $\hat{\sigma}_b^2$. This is supported by the simulation results in Shao (1988*a*). When a_m is chosen to be large, $\hat{\sigma}_a^2$ is asymptotically equivalent to $\hat{\sigma}_b^2$ in some cases (Example 2.2). The consistency of $\hat{\sigma}_a^2$ is studied in Section 3.

Example 2.1. Let $\hat{\theta}_n$ be the sample p-quantile with $0 < p < 1$ and $\hat{\theta}_n^*$ be the bootstrap sample p-quantile ($m = n$). Suppose that the derivative of $F(x)$ is continuous in a neighborhood of $F^{-1}(p)$. Let $\Delta_n^* = \Delta_{nm}^*(a_m)$ (2.3) with $m = n$ and $a_m \equiv 1$. Then

(2.4) $$n \operatorname{Var}_* \Delta_n^* \to \sigma^2 \quad \text{a.s.}$$

The counter-example in Ghosh *et al.* (1984) shows that some tail condition is needed on F to ensure the consistency of the bootstrap estimator $\hat{\sigma}_b^2 = n \operatorname{Var}_* \hat{\theta}_n^*$. Under a moment condition $E|X_1|^{\varepsilon} < \infty$ for some $\varepsilon > 0$, $\hat{\sigma}_b^2 \to \sigma^2$ a.s. is proved in the same article. On the other hand, the consistency of the modified bootstrap estimator $n \operatorname{Var}_* \Delta_n^*$ requires no moment condition. The truncating in (2.3) circumvents the problems of the bootstrap with heavy-tailed distribution.

To show (2.4), first note that from Proposition 5.1 of Bickel and Freedman (1981),

$$n^{1/2}(\hat{\theta}_{nm}^* - \hat{\theta}_n) \underset{d^*}{\to} Z \quad \text{a.s.} ,$$

where Z is a random variable distributed as $N(0, \sigma^2)$ and $\underset{d^*}{\to}$ denotes convergence in the conditional distribution. This implies

$$n^{1/2}\Delta_n^* \underset{d^*}{\rightarrow} Z \qquad \text{a.s.}$$

Then (2.4) is implied by

$$(2.5)\quad (2+\delta)^{-1}E_*|n^{1/2}\Delta_n^*|^{2+\delta} = \int_0^\infty t^{1+\delta}P_*(n^{1/2}|\Delta_n^*| > t)dt = O(1) \qquad \text{a.s.}$$

for some $\delta > 0$. From $|\Delta_n^*| \le 1$, the left-hand side of (2.5) is equal to

$$\int_0^{n^{1/2}} t^{1+\delta}P_*(n^{1/2}|\Delta_n^*| > t)dt \le \int_0^{n^{1/2}} t^{1+\delta}P_*(n^{1/2}|\hat{\theta}_{nm}^* - \hat{\theta}_n| > t)dt\,.$$

The rest of the proof is the same as that in the proof of Theorem 1 of Ghosh *et al.* (1984).

Example 2.2. Let $\hat{\theta}_n$ be an L-estimator

$$(2.6)\qquad \sum_{i=1}^n c_{ni}X_{(i)} \quad \text{with} \quad \sup_n \sum_{i=1}^n |c_{ni}| = \rho < \infty\,,$$

where $X_{(i)}$ is the i-th order statistic of $X_1,\ldots,X_n$. Note that the sample mean and the sample p-quantile $(0 < p < 1)$ are special cases of (2.6).

Under the condition that $E|X_1|^\varepsilon < \infty$ for some $\varepsilon > 0$, if $\Delta_{nm}^*(a_m)$ is defined as in (2.3) with $a_m \ge cn^{1/\varepsilon}$, where c is a positive constant, then

$$P(\hat{\sigma}_a^2 = \hat{\sigma}_b^2 \text{ for all sufficiently large } n) = 1\,,$$

since

$$|\hat{\theta}_{nm}^* - \hat{\theta}_n|/a_m \le 2\rho(|X_{(n)}| + |X_{(1)}|)/cn^{1/\varepsilon} \to 0 \qquad \text{a.s.}$$

by Lemma 3 of Ghosh *et al.* (1984).

A similar result can be obtained if $\hat{\theta}_n$ is a U- or V-statistic.

The following analog of Helly's theorem (Serfling (1980), p. 352) will be used in the proof of Theorem 2.1. Its proof can be found in Shao (1988*b*).

LEMMA 2.1. *Let Y_n be random n-vectors, $n = 1, 2, \ldots$, and Z be a random variable. Suppose that for any fixed n and given $Y_n = y$, $Z_n(y)$ is a random variable and as $n \to \infty$,*

$$(2.7)\qquad P(Z_n(Y_n) \le t\,|\,Y_n) \to P(Z \le t) \quad \textit{in probability}$$

for any t at which $P(Z \leq t)$ is continuous. Then as $n \to \infty$,

$$E[h(Z_n(Y_n)) \mid Y_n] \to E[h(Z)] \quad \textit{in probability}$$

for any real-valued bounded continuous function h.

THEOREM 2.1. *Suppose that as n and $m \to \infty$,*

$$m^{1/2}(\hat{\theta}^*_{nm} - \hat{\theta}_n) \underset{d^*}{\to} Z \quad \textit{in probability} \tag{2.8}$$

in the sense of (2.7) *with $Y_n = (X_1, \ldots, X_n)$ and $Z_n(Y_n) = m^{1/2}(\hat{\theta}^*_{nm} - \hat{\theta}_n)$, where Z is a random variable with mean zero and variance σ^2. Let $\Delta^*_{nm}(a_m^{(k)})$ be defined as in* (2.3) *with $a_m^{(2)} \geq a_m^{(1)}$ and $m^{1/2}a_m^{(k)} \to \infty$, $k = 1, 2$. Then*

$$m \operatorname{Var}_* [\Delta^*_{nm}(a_m^{(2)})] \to \sigma^2 \quad \textit{in probability} \tag{2.9}$$

implies

$$m \operatorname{Var}_* [\Delta^*_{nm}(a_m^{(1)})] \to \sigma^2 \quad \textit{in probability}\,. \tag{2.10}$$

Remarks. (i) If in (2.8)–(2.10), the statement "in probability" is replaced by "a.s.", then the result follows immediately from Lemma 1.4B and Theorem 1.4A of Serfling (1980).

(ii) The result shows that the consistency of $\hat{\sigma}_b^2$ implies the consistency of $\hat{\sigma}_a^2$ (by taking $a_m^{(2)} \equiv \infty$).

PROOF. Denote $m^{1/2}\Delta^*_{nm}(a_m^{(k)})$ by Z^{*k}_{nm}, $k = 1, 2$. Condition (2.8) and $m^{1/2}a_m^{(k)} \to \infty$ imply

$$Z^{*k}_{nm} \underset{d^*}{\to} Z \quad \text{in probability}, \quad k = 1, 2\,,$$

which implies (Ghosh *et al.* (1984), Lemma 2) that (2.9) and (2.10) are equivalent respectively to

$$E_*(Z^{*2}_{nm})^2 \to \sigma^2 \quad \text{in probability}$$

and

$$E_*(Z^{*1}_{nm})^2 \to \sigma^2 \quad \text{in probability}\,.$$

By Lemma 2.1, as n and $m \to \infty$,

$$E_*[(Z^{*k}_{nm})^2 I((Z^{*k}_{nm})^2 \leq M)] - E[Z^2 I(Z^2 \leq M)] \to 0 \quad \text{in probability}\,, \tag{2.11}$$

$$k = 1, 2\,,$$

where $I(A)$ is the indicator function of a set A.

If (2.10) does not hold, then there is a set $\Omega = \{n_l, m_l, l = 1, 2, \dots\}$ with infinitely many elements such that

$$\lim_{l \to \infty} P(|E_*(Z^{*1}_{n_l m_l})^2 - \sigma^2| > \delta) \neq 0 \tag{2.12}$$

for some $\delta > 0$. For any $\varepsilon > 0$, choose an $M > 0$ such that $E[Z^2 I(Z^2 > M)] < \varepsilon$. From (2.9) and (2.11), there is a subset $\{n_j, m_j, j = 1, 2, \dots\}$ of Ω such that as $j \to \infty$,

$$E_*(Z^{*2}_{n_j m_j})^2 \to \sigma^2 \quad \text{a.s.}\,,$$

and

$$E_*[(Z^{*k}_{n_j m_j})^2 I((Z^{*k}_{n_j m_j})^2 \leq M)] - E[Z^2 I(Z^2 \leq M)] \to 0 \quad \text{a.s.}, \quad k = 1, 2\,.$$

Hence for almost all $X_1, X_2, \dots$, there is an $N > 0$ such that when $j > N$,

$$E_*[(Z^{*2}_{n_j m_j})^2 I((Z^{*2}_{n_j m_j})^2 \leq M)] \geq E[Z^2 I(Z^2 \leq M)] - \varepsilon$$

and

$$E_*(Z^{*2}_{n_j m_j})^2 \leq \sigma^2 + \varepsilon\,,$$

and therefore

$$\begin{aligned} E_*[(Z^{*1}_{n_j m_j})^2 I((Z^{*1}_{n_j m_j})^2 > M)] &\leq E_*[(Z^{*2}_{n_j m_j})^2 I((Z^{*2}_{n_j m_j})^2 > M)] \\ &\leq \sigma^2 - E[Z^2 I(Z^2 \leq M)] + 2\varepsilon \\ &= E[Z^2 I(Z^2 > M)] + 2\varepsilon \leq 3\varepsilon\,. \end{aligned}$$

Then

$$\begin{aligned} &|E_*(Z^{*1}_{n_j m_j})^2 - \sigma^2| \\ &\quad \leq E_*[(Z^{*1}_{n_j m_j})^2 I((Z^{*1}_{n_j m_j})^2 > M)] + E[Z^2 I(Z^2 > M)] \\ &\qquad + |E_*[(Z^{*1}_{n_j m_j})^2 I((Z^{*1}_{n_j m_j})^2 \leq M)] - E[Z^2 I(Z^2 \leq M)]| \\ &\quad \leq 4\varepsilon + |E_*[(Z^{*1}_{n_j m_j})^2 I((Z^{*1}_{n_j m_j})^2 \leq M)] - E[Z^2 I(Z^2 \leq M)]|\,. \end{aligned}$$

Letting $j \to \infty$ and $\varepsilon \to \infty$, we have

$$E_*(Z^{*1}_{n,m_j})^2 \to \sigma^2 \qquad \text{a.s.}$$

This contradicts (2.12) and thus the result. □

3. Consistency of the modified bootstrap variance estimator

For the asymptotic validity of the bootstrap methods, we require some smoothness conditions for the functional T. The following two definitions of differentiability of T can be found, respectively, in Serfling (1980) and Parr (1985).

DEFINITION 3.1. (i) A functional T is said to be Fréchet differentiable at F with respect to a norm $\| \ \|$ on Ξ iff there exits a real-valued function ϕ depending only on T and F such that $\int \phi(x)dF(x) = 0$ and

$$\left| T(G) - T(F) - \int \phi(x)d[G-F](x) \right| \Big/ \|G-F\| \to 0 \tag{3.1}$$

as $\|G-F\| \to 0$, for $G \in \Xi$.

(ii) A functional T is said to be strongly Fréchet differentiable at F with respect to $\| \ \|$ iff (3.1) holds with F replaced by G_1 as $\|G-F\| + \|G_1 - F\| \to 0$, for $G, G_1 \in \Xi$.

The function ϕ is the usual influence function of T (Hampel (1974)). Obviously, strong Fréchet differentiability implies Fréchet differentiability. Examples of Fréchet differentiable or strongly Fréchet differentiable T, including certain types of V-statistics and M- and L-estimators, can be found in Serfling (1980) and Parr (1985). We prove the consistency of $\hat{\sigma}_a^2$ for Fréchet differentiable and strongly Fréchet differentiable T in Theorems 3.1 and 3.2, respectively.

We will assume in the sequel that the influence function ϕ satisfies

$$0 < \int \phi^2(x)dF(x) = \sigma^2 < \infty .$$

If T is Fréchet differentiable, then

$$\begin{aligned} \hat{\theta}_n &= \theta + n^{-1} \sum_{i=1}^{n} \phi(X_i) + R_n , \\ \hat{\theta}^*_{nm} &= \theta + m^{-1} \sum_{i=1}^{m} \phi(X_i^*) + R^*_{nm} \end{aligned} \tag{3.2}$$

and

$$\hat{\theta}^*_{nm} - \hat{\theta}_n = m^{-1} \sum_{i=1}^{m} \phi(X_i^*) - n^{-1} \sum_{i=1}^{n} \phi(X_i) + U^*_{nm}\,, \tag{3.3}$$

where R_n, R^*_{nm} and $U^*_{nm} = R^*_{nm} - R_n$ are the remainders. From Serfling ((1980), p. 218), $R_n = o_p(n^{-1/2})$, which implies

$$n^{1/2}(\hat{\theta}_n - \theta) \underset{d}{\rightarrow} Z\,, \tag{3.4}$$

where Z is distributed as $N(0, \sigma^2)$. Finding the orders of R^*_{nm} and U^*_{nm} is crucial for the asymptotic analysis based on the bootstrap method. It is shown in Proposition 3.2 that if T is Fréchet differentiable (or strongly Fréchet differentiable), then $U^*_{nm} = o_p(m^{-1/2})$ in probability (or a.s.), which implies

$$m^{1/2}(\hat{\theta}^*_{nm} - \hat{\theta}_n) \underset{d^*}{\rightarrow} Z \quad \text{in probability} \qquad \text{(or a.s.)}\,. \tag{3.5}$$

That is, the bootstrap approximation to the distribution of $n^{1/2}(\hat{\theta}_n - \theta)$ is consistent. (3.4) and (3.5) still hold if T has weak versions of differential (see Gill (1989)).

For simplicity, from now on we use $\|\ \|$ to denote the sup norm of a bounded real-valued function h, i.e., $\|h\| = \sup_{|t|<\infty} |h(t)|$.

THEOREM 3.1. *Let $\hat{\sigma}^2_a$ be defined as in* (2.2) *with a_m satisfying*

$$m^{1/2} a_m \rightarrow \infty \qquad \textit{and} \qquad a_m \le e^{m^p} \tag{3.6}$$

for some $0 < p < 1$. Assume that T is Fréchet differentiable with respect to $\|\ \|$ and $(\log (n))^q \le m \le \lambda n$, where $\lambda > 0$ is a constant and $q = 1/p$. Then as n and $m \rightarrow \infty$,

$$\hat{\sigma}^2_a \rightarrow \sigma^2 \qquad \textit{in probability}\,. \tag{3.7}$$

THEOREM 3.2. *Let $\hat{\sigma}^2_a$ be defined as in* (2.2) *with a_m satisfying* (3.6). *Assume that T is strongly Fréchet differentiable with respect to $\|\ \|$ and $m^p \ge \log (n)$. Then as n and $m \rightarrow \infty$,*

$$\hat{\sigma}^2_a \rightarrow \sigma^2 \qquad \textit{a.s.} \tag{3.8}$$

We first prove the following results.

PROPOSITION 3.1. (i) *Let $\varepsilon > 0$ be given. Then there is a constant $\rho > 0$ such that for all n and m,* (3.9) $$P_*(\|F^*_{nm} - F_n\| > \varepsilon) \le \rho e^{-2m\varepsilon^2}.$$

(ii) *If $m = m(n)$ satisfies*

(3.10) $$\sum_{n=1}^{\infty} e^{-2m(n)\varepsilon^2} < \infty$$

for any $\varepsilon > 0$, then as $n \to \infty$, for almost all $X_1, X_2, \ldots,$

$$\|F^*_{nm} - F\| \to 0 \quad a.s. \quad P_*.$$

(iii) *Let $\Delta^*_{nm} = \Delta^*_{nm}(a_m)$ (2.3) with $a_m \le e^{m^p}$ for some $0 < p < 1$ and*

(3.11) $$B_{nm} = \{\|F^*_{nm} - F\| > \eta\}$$

for a constant $\eta > 0$. Then for any $\alpha \ge 1$, as n and $m \to \infty$,

$$E_*|m^{1/2}\Delta^*_{nm} I(B_{nm})|^{\alpha} \to 0 \quad a.s.$$

Remark. A consequence of (3.9) is that for any $\varepsilon > 0$,

(3.12) $$P_*(\|F^*_{nm} - F\| > \varepsilon) \to 0 \quad \text{in probability},$$

which is equivalent to the result in Corollary 4.1 of Bickel and Freedman (1981). The result in (ii) is stronger than (3.12) but m needs to be a function of n and satisfies (3.10). Note that (3.10) holds if $m(n) \ge k_n \log(n)$ with any sequence k_n satisfying $k_n \to \infty$ as $n \to \infty$.

PROOF. Equation (3.9) follows from the inequality of Dvoretzky, Kiefer and Wolfowitz (see Serfling (1980), pp. 59–60). The result in (ii) is a direct consequence of (3.9) and (3.10), the Borel-Cantelli Lemma, and $\|F_n - F\| \to 0$ a.s. The result in (iii) follows from (3.9) and $\|F_n - F\| \to 0$ a.s. □

PROPOSITION 3.2. (i) *Suppose that T is Fréchet differentiable. Assume also $m \le \lambda n$ for a constant $\lambda > 0$. Let R^*_{nm} be defined as in* (3.2). *Then as n and $m \to \infty$,*

$$m^{1/2} R^*_{nm} \underset{d^*}{\to} 0 \quad \textit{in probability}.$$

(ii) *Suppose that T is strongly Fréchet differentiable. Let U^*_{nm} be defined as in* (3.3). *Then for any $\tau > 0$, as n and $m \to \infty$,*

$$P_*(|m^{1/2} U^*_{nm}| > \tau) \to 0 \quad a.s.$$

PROOF. (i) Let $\tau > 0$ be given. By the Fréchet differentiability of T, for any $\varepsilon > 0$, there is a $\delta_\varepsilon > 0$ such that

$$|R^*_{nm}| < \varepsilon \| F^*_{nm} - F \|$$

whenever $\| F^*_{nm} - F \| < \delta_\varepsilon$. Then

$$P_*(m^{1/2} | R^*_{nm}| > \tau) \leq P_*(m^{1/2} \| F^*_{nm} - F \| > \tau / \varepsilon) + P_*(\| F^*_{nm} - F \| > \delta_\varepsilon) . \tag{3.13}$$

Let

$$A_n = \{(\lambda n)^{1/2} \| F_n - F \| \leq \tau / 2\varepsilon\}$$

and A^c_n and $I(A_n)$ be the complement and the indicator function of A_n, respectively. From $m^{1/2} \| F^*_{nm} - F \| \leq m^{1/2} \| F^*_{nm} - F_n \| + (\lambda n)^{1/2} \| F_n - F \|$, we have

$$\begin{aligned} E[P_*(m^{1/2} \| F^*_{nm} - F \| > \tau / \varepsilon)] &\leq E[I(A_n) P_*(m^{1/2} \| F^*_{nm} - F \| > \tau / \varepsilon)] + P(A^c_n) \\ &\leq E[P_*(m^{1/2} \| F^*_{nm} - F_n \| > \tau / 2\varepsilon)] + P(A^c_n) \\ &\leq \rho [e^{-(1/2)(\tau/\varepsilon)^2} + e^{-(1/2)\lambda^{-1}(\tau/\varepsilon)^2}] , \end{aligned}$$

by (3.9) and applying the Dvoretzky-Kiefer-Wolfowitz inequality to $P(A^c_n)$. From (3.12) and (3.13), the result in (i) follows since ε is arbitrary.

(ii) By the strong Fréchet differentiability of T, U^*_{nm} satisfies the following property: for any $\varepsilon > 0$, there is a $\delta_\varepsilon > 0$ such that

$$| U^*_{nm}| < \varepsilon \| F^*_{nm} - F_n \|$$

whenever $\| F^*_{nm} - F \| + \| F_n - F \| < \delta_\varepsilon$. For almost all $\omega = (X_1, X_2, \ldots)$, there is an $N(\omega) > 0$ such that $\| F_n - F \| < \delta_\varepsilon / 4$ for $n > N(\omega)$. Then for this ω and a given $\tau > 0$,

$$\begin{aligned} P_*(| m^{1/2} U^*_{nm}| > \tau) &\leq P_*(m^{1/2} \| F^*_{nm} - F_n \| > \tau / \varepsilon) + P_*(\| F^*_{nm} - F_n \| > \delta_\varepsilon / 2) \\ &\leq \rho [e^{-2(\tau/\varepsilon)^2} + e^{-(1/2) m \delta_\varepsilon^2}] . \end{aligned}$$

The result in (ii) follows since $e^{-(1/2) m \delta_\varepsilon^2} \to 0$ as $m \to \infty$. □

PROOF OF THEOREM 3.1. From Theorem 2.1, we need only to show (3.7) for $\hat{\sigma}^2_a$ with $a_m = e^{m^p}$. Let $\Delta^*_{nm} = \Delta^*_{nm}(e^{m^p})$. By the Fréchet differentiability of T, there is an $\eta > 0$ such that $|R^*_{nm}| < \|F^*_{nm} - F\|$ whenever $\|F^*_{nm} - F\| < \eta$. From Proposition 3.1(iii), (3.7) is implied by

$$m \operatorname{Var}_* [\Delta^*_{nm} I(B^c_{nm})] \to \sigma^2 \quad \text{in probability}, \tag{3.14}$$

where B^c_{nm} is the complement of B_{nm} defined in (3.11). Let

$$D_{nm} = \left\{ \max_{i \le n} |\phi(X_i)| \le 2^{-1} e^{m^p} - 1 \right\}$$

and

$$G_n = \{|R_n| > 1\}.$$

From $E\phi^2(X_1) < \infty$, $\max_{i \le n} |\phi(X_i)|/n^{1/2} \to 0$ a.s. (Ghosh *et al.* (1984), Lemma 3), which implies $\max_{i \le n} |\phi(X_i)|/e^{m^p} \to 0$ a.s. since $m^p \ge \log(n)$ is assumed. Then $P(D_{nm}) + P(G_n) \to 0$ as n and $m \to \infty$. Let $(X_1, \ldots, X_n) \in D^c_{nm} \cap G^c_n$. If $\|F^*_{nm} - F\| \le \eta$, then

$$\begin{aligned} |\hat\theta^*_{nm} - \hat\theta_n| &\le m^{-1} \sum_{i=1}^m |\phi(X_i^*)| + n^{-1} \sum_{i=1}^n |\phi(X_i)| + |R^*_{nm}| + |R_n| \\ &\le 2 \max_{i \le n} |\phi(X_i)| + 2 \le e^{m^p}, \end{aligned}$$

and therefore

$$m \operatorname{Var}_* [\Delta^*_{nm} I(B^c_{nm})] = m \operatorname{Var}_* [(\hat\theta^*_{nm} - \hat\theta_n) I(B^c_{nm})].$$

Then (3.14) is equivalent to

$$m \operatorname{Var}_* [(\hat\theta^*_{nm} - \hat\theta_n) I(B^c_{nm})] \to \sigma^2 \quad \text{in probability}.$$

From (3.3), it suffices to show that

$$m \operatorname{Var}_* \left[\left(m^{-1} \sum_{i=1}^m \phi(X_i^*) - n^{-1} \sum_{i=1}^n \phi(X_i) \right) I(B^c_{nm}) \right] \to \sigma^2 \quad \text{a.s.} \tag{3.15}$$

and

$$E_*[m(R^*_{nm} - R_n)^2 I(B^c_{nm})] \to 0 \quad \text{in probability}. \tag{3.16}$$

By the strong law of large numbers,

$$m\,\mathrm{Var}_*\left[m^{-1}\sum_{i=1}^{m}\phi(X_i^*)-n^{-1}\sum_{i=1}^{n}\phi(X_i)\right]$$
$$=n^{-1}\sum_{i=1}^{n}\left[\phi(X_i)-n^{-1}\sum_{i=1}^{n}\phi(X_i)\right]^2\to\sigma^2\qquad\text{a.s.}$$

Note that if $(X_1,\ldots,X_n)\in D_{nm}$, $\left|m^{-1}\sum_{i=1}^{m}\phi(X_i^*)-n^{-1}\sum_{i=1}^{n}\phi(X_i)\right|\le e^{m^p}$. Hence from Proposition 3.1(iii),

$$m\,\mathrm{Var}_*\left[\left(m^{-1}\sum_{i=1}^{m}\phi(X_i^*)-n^{-1}\sum_{i=1}^{n}\phi(X_i)\right)I(B_{nm})\right]\to 0\qquad\text{a.s.}$$

Thus (3.15) holds. Note that $mR_n^2\le\lambda nR_n^2=o_p(1)$. Hence (3.16) is implied by

$$(3.17)\qquad E_*[(m^{1/2}R_{nm}^*)^2I(B_{nm}^c)]\to 0\qquad\text{in probability}\,.$$

For an arbitrary $\varepsilon>0$, there is an $M>0$ such that

$$P(A_n)\le\varepsilon$$

for all n, where $A_n=\{n^{1/2}\|F_n-F\|>M\}$. Let $\delta>0$ be fixed. For any fixed n, m and $(X_1,\ldots,X_n)\in A_n^c$,

$$\begin{aligned}
&E_*[|m^{1/2}R_{nm}^*|^{2+\delta}I(B_{nm}^c)]\\
&\quad=(2+\delta)\int_0^\infty t^{1+\delta}P_*(|m^{1/2}R_{nm}^*I(B_{nm}^c)|>t)dt\\
&\quad\le(2+\delta)\int_0^\infty t^{1+\delta}P_*(m^{1/2}\|F_{nm}^*-F\|>t)dt\\
&\quad\le(2+\delta)\int_0^\infty t^{1+\delta}P_*(m^{1/2}\|F_{nm}^*-F_n\|+(\lambda n)^{1/2}\|F_n-F\|>t)dt\\
&\quad\le(2+\delta)\int_0^\infty t^{1+\delta}P_*(m^{1/2}\|F_{nm}^*-F_n\|>t-\lambda^{1/2}M)dt\\
&\quad=(2+\delta)\left[\int_0^{\lambda^{1/2}M}t^{1+\delta}P_*(m^{1/2}\|F_{nm}^*-F_n\|>t-\lambda^{1/2}M)dt\right.\\
&\qquad\qquad\left.+\int_{\lambda^{1/2}M}^\infty t^{1+\delta}P_*(m^{1/2}\|F_{nm}^*-F_n\|>t-\lambda^{1/2}M)dt\right]\\
&\quad\le(\lambda^{1/2}M)^{2+\delta}+\rho(2+\delta)\int_{\lambda^{1/2}M}^\infty t^{1+\delta}e^{-2(t-\lambda^{1/2}M)^2}dt\,,
\end{aligned}$$

which is equal to some finite constant (independent of n and m), say $C(M)$, where the last inequality follows from (3.9). This shows that

$$A_n^c \subset \{E_*[|m^{1/2}R_{nm}^*|^{2+\delta}I(B_{nm}^c)] \le C(M)\}$$

and therefore

$$P(E_*[|m^{1/2}R_{nm}^*|^{2+\delta}I(B_{nm}^c)] > C(M)) \le P(A_n) \le \varepsilon\,. \tag{3.18}$$

Let $\tau > 0$ be given. Choose an $\alpha > \tau$ such that

$$C(M) < \alpha^{(1/2)\delta}\tau/4\,. \tag{3.19}$$

Let $\zeta_{nm}^* = (m^{1/2}R_{nm}^*)^2 I(B_{nm}^c)$. From

$$\begin{aligned} E_*\zeta_{nm}^* &\le \tau/2 + E_*[\zeta_{nm}^* I(\zeta_{nm}^* > \tau/2)] \\ &\le \tau/2 + \alpha P_*((m^{1/2}R_{nm}^*)^2 > \tau/2) + E_*[\zeta_{nm}^* I(\zeta_{nm}^* > \alpha)] \\ &\le \tau/2 + \alpha P_*((m^{1/2}R_{nm}^*)^2 > \tau/2) + \alpha^{-(1/2)\delta}E_*[|m^{1/2}R_{nm}^*|^{2+\delta}I(B_{nm}^c)]\,, \end{aligned}$$

and (3.18)–(3.19), we have

$$P(E_*[(m^{1/2}R_{nm}^*)^2 I(B_{nm}^c)] > \tau) \le P[P_*((m^{1/2}R_{nm}^*)^2 > \tau/2) > \tau/4\alpha] + \varepsilon\,.$$

By Proposition 3.2(i), $\lim_{n,m\to\infty} P[P_*((m^{1/2}R_{nm}^*)^2 > \tau/2) > \tau/4\alpha] = 0$. Hence

$$\limsup_{n,m\to\infty} P(E_*[(m^{1/2}R_{nm}^*)^2 I(B_{nm}^c)] > \tau) \le \varepsilon\,.$$

(3.17) follows since ε is arbitrary. □

PROOF OF THEOREM 3.2. It suffices to show (3.8) for $\hat{\sigma}_a^2$ with $\Delta_{nm}^* = \Delta_{nm}^*(e^{m^p})$. By the strong Fréchet differentiability of T, there exists an $\eta > 0$ such that when $\|F_{nm}^* - F\| + \|F_n - F\| < \eta$, $|U_{nm}^*| < \|F_{nm}^* - F_n\|$. Let $B_{nm} = \{\|F_{nm}^* - F\| > \eta/2\}$. Then from Proposition 3.1(iii), we need only to show that

$$m\,\mathrm{Var}_*\,[\Delta_{nm}^* I(B_{nm}^c)] \to \sigma^2 \qquad \text{a.s.}$$

Note that $\|F_n - F\| \to 0$ and $\max_{i\le n} |\phi(X_i)|/e^{m^p} \to 0$ a.s. Then for almost all $\omega = (X_1, X_2, \dots)$, there is an $N(\omega) > 0$ such that for all $n > N(\omega)$ and $m > N(\omega)$,

$$\max_{i\le n} |\phi(X_i)| \le 2^{-1}e^{m^p} - 1$$

and

(3.20) $$\|F_n - F\| < \eta/2\,.$$

For this ω, using the same argument as in the proof of Theorem 3.1, one can show that

$$m\ \mathrm{Var}_*\,[\Delta^*_{nm}I(B^c_{nm})] = m\ \mathrm{Var}_*\,[(\hat{\theta}^*_{nm} - \hat{\theta}_n)I(B^c_{nm})]\,.$$

Hence by (3.15), what remains to be shown is that

$$E_*[(m^{1/2}U^*_{nm})^2 I(B^c_{nm})] \to 0 \qquad \text{a.s.}$$

Let $\delta > 0$, and $\varepsilon > 0$, $\alpha > 0$ be arbitrary. From

$$E_*[(m^{1/2}U^*_{nm})^2 I(B^c_{nm})] \le \varepsilon + \alpha P_*((m^{1/2}U^*_{nm})^2 > \varepsilon) + \alpha^{-(1/2)\delta} E_*[|m^{1/2}U^*_{nm}|^{2+\delta} I(B^c_{nm})]\,,$$

it suffices to show that

(3.21) $$E_*[|m^{1/2}U^*_{nm}|^{2+\delta} I(B^c_{nm})] \le c < \infty$$

for all n, $m > N(\omega)$, since from Proposition 3.2(ii), $P_*((m^{1/2}U^*_{nm})^2 > \varepsilon) \to 0$ a.s., and ε and α are arbitrary. From (3.20), $|U^*_{nm}I(B^c_{nm})| \le \|F^*_{nm} - F_n\|$. Hence (3.21) follows from

$$\begin{aligned} E_*[|m^{1/2}U^*_{nm}|^{2+\delta} I(B^c_{nm})] &= (2+\delta)\int_0^\infty t^{1+\delta} P_*(|m^{1/2}U^*_{nm}I(B^c_{nm})| > t)dt \\ &\le (2+\delta)\int_0^\infty t^{1+\delta} P_*(m^{1/2}\|F^*_{nm} - F_n\| > t)dt \\ &\le \rho(2+\delta)\int_0^\infty t^{1+\delta} e^{-2t^2} dt\,, \end{aligned}$$

where the last inequality follows from (3.9). This proves the theorem. □

4. Discussions

(i) *The choice of* a_m. The sequence a_m in (2.3) can be chosen to be a function of $X_1, \ldots, X_n$. Apparently, all the results in Section 3 still hold as long as (3.6) is satisfied for almost all $X_1, X_2, \ldots$. A reasonable choice is $a_m = \max(\rho|\hat{\theta}_n|, c)$, where ρ and c are positive constants. With this choice, $|\hat{\theta}^*_{nm} - \hat{\theta}_n|$ is truncated if it exceeds $100\rho\%|\hat{\theta}_n|$.

(ii) *Comparison to variance estimators based on bootstrap quantiles.* From (3.5), another consistent estimator for σ^2 is $\hat{\sigma}^2_q = (q^*_{1-t} - q^*_t)^2/(z_{1-t} - z_t)^2$, where q^*_t and z_t are the t-quantiles of the bootstrap distribution and the standard normal distribution, respectively. However, the modified bootstrap estimator $\hat{\sigma}_a^2$ is much easier and cheaper to compute than $\hat{\sigma}_q^2$ (note that the bootstrap estimators usually require Monte Carlo approximations). According to Efron (1987), the Monte Carlo approximation of $\hat{\sigma}_a^2$ usually requires $B = 100 \sim 200$ bootstrap replications (see (2.1)), whereas the Monte Carlo approximation of the bootstrap quantile q_t^* is more costly, requiring $1000 \sim 2000$ bootstrap replications.

(iii) *The differentiability assumption.* In Section 3, consistency of $\hat{\sigma}_a^2$ was proved under the assumption that T is Fréchet differentiable. There are some statistical functionals that are not Fréchet differentiable (or the Fréchet differentiability is not known). In fact, results (3.4) and (3.5) (the weak consistency of the bootstrap distribution) may hold under weaker assumptions on T (e.g., Gill (1989)). However, the consistency of the bootstrap variance estimators $\hat{\sigma}_b^2$ and $\hat{\sigma}_a^2$ for general functionals is still an unresolved problem.

(iv) *Extension to the multivariate case.* Let $T = (T_1, \ldots, T_k)$, where T_i's are functionals from Ξ to the real line. Definition 3.1 in Section 3 can be extended to T and $\hat{\sigma}_a^2$ (2.2) has the obvious extension with $\hat{\theta}_n = T(F_n)$. Proofs of the consistency of $\hat{\sigma}_a^2$ are the same as before.

Acknowledgement

The author would like to thank the referee for helpful comments.

REFERENCES

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. (1987). Better bootstrap confidence intervals, *J. Amer. Statist. Assoc.*, **82**, 171–200.

Ghosh, M., Parr, W. C., Singh, K. and Babu, G. J. (1984). A note on bootstrapping the sample median, *Ann. Statist.*, **12**, 1130–1135.

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part I), *Scand. J. Statist.*, **16**, 97–128.

Hampel, F. R. (1974). The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.*, **69**, 383–397.

Parr, W. C. (1985). Jackknifing differentiable statistical functionals, *J. Roy. Statist. Soc. Ser. B*, **47**, 56–66.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data, *J. Amer. Statist. Assoc.*, **83**, 231–241.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Shao, J. (1988*a*). A note on bootstrap variance estimation, Tech. Report, Dept. of Statistics, Purdue University.

Shao, J. (1988*b*). Consistency of jackknife estimators of the variances of sample quantiles, *Comm. Statist. A—Theory Methods*, **17**, 3017–3028.

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

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works, *Comm. Statist. A—Theory Methods*, **15**, 3193–3203.