Vietnam Journal of Mathematics **26**:4 (1998) 351–356

Vietnam Journal of MATHEMATICS

Short Communication

On the Asymptotic Accuracy of the Bootstrap with Random Sample Size*

Nguyen Van Toan
Department of Mathematics, College of Science
Hue University, Hue, Vietnam

Received June 9, 1997
Revised June 11, 1998

1. Introduction

Let $X_1, X_2, \dots$ be independent random variables with the common distribution function F. For any but fixed $n \in \mathbf{N}$, denote by F_n the empirical distribution of $(X_1, \dots, X_n)$ and by $X^*_{n1}, X^*_{n2}, \dots$ independent identically distributed (i.i.d.) random variables with the distribution F_n. By N_n, we mean a positive integer-valued random variable independent of $X_1, \dots, X_n$ such that

$$N_n \to_p \infty \text{ as } n \to \infty, \tag{1}$$

where $\to_p$ denotes the convergence in probability.

We study the following bootstrap procedure with a random sample size for estimating $P(\sqrt{n}\,(\bar{X}_n - \mu) < x)$, where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, $\mu = E(X_1)$ is the expectation of X_1. Then the bootstrap estimate is $P^*(\sqrt{n}(\bar{X}^*_n - \bar{X}_n) < x)$ and the bootstrap estimate with random sample size N_n will be $P^*(\sqrt{n}(\bar{X}^{*N_n}_n - \frac{N_n}{n}\bar{X}_n) < x)$ or $P^*(\sqrt{N_n}(\bar{X}^*_{N_n} - \bar{X}_n) < x)$, where P^* denotes the conditional law $P(\dots|X_1, \dots, X_n)$, $\bar{X}^*_n = n^{-1}\sum_{n=1}^n X^*_{ni}$, $\bar{X}^{*N_n}_n = n^{-1}\sum_{i=1}^{N_n} X^*_{ni}$ and $\bar{X}^*_{N_n} = N_n^{-1}\sum_{i=1}^{N_n} X^*_{ni}$. It is known that bootstrap is (weakly) consistent if and only if X_1 belongs to the domain of attraction of the normal law (see [1–5]) and then if

$$\frac{N_n}{n} \to_p 1 \text{ as } n \to \infty,$$

$P^*(\sqrt{n}(\bar{X}^{*N_n}_n - \frac{N_n}{n}\bar{X}_n) < x)$ can be used as an estimate of $P(\sqrt{n}(\bar{X}_n - \mu) < x)$ (see [7, 8, 10, 11]). In this case, when $EX_1^2 < \infty$ and (1) holds, $P^*(\sqrt{N_n}(\bar{X}^*_{N_n} - \bar{X}_n) < x)$ can be used as given in [10].

* This research is supported by The Fundamental Research Program 1998–1999.

The purpose of this paper is to study the rate of convergence of the bootstrap estimates with a random sample size in that case.

2. Results

In what follows, set $s_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ and denote by $\sigma^2 = D(X_1)$ the variance of X_1. Let $\| \ldots \|_\infty = \sup_{-\infty<x<\infty} | \ldots |$.

Our main results are presented in the next two theorems, namely, in Theorem 1, we first study the uniform convergence to zero of the discrepancy between the actual distribution of $\sqrt{n}\,(\overline{X}_n - \mu)$ and the approximation $\sqrt{n}\,(\overline{X}_n^{*N_n} - \frac{N_n}{n}\overline{X}_n)$ of it (Part A) and then we study the uniform and non-uniform convergence to zero of the discrepancy between the actual distribution of $\frac{\sqrt{n}}{\sigma}(\overline{X}_n - \mu)$ and the approximation $\frac{\sqrt{n}}{s_n}(\overline{X}_n^{*N_n} - \frac{N_n}{n}\overline{X}_n)$ of it (Part B).

Theorem 1. *Let $X_1, X_2, \ldots$ be i.i.d. variables with distribution F. Let N_n be a positive integer valued random variable independent of $X_1, X_2, \ldots$. Let F_n be the empirical distribution of $X_1, \ldots, X_n$. Given $X_1, \ldots, X_n$, let $X_{n1}^*, X_{n2}^*, \ldots$ be conditionally independent, with common distribution F_n.*

(A) *If $EX_1^4 < \infty$, $EN_n = n + O(\sqrt{n \log\log n})$ and $DN_n = O(n \log\log n)$, then*

$$\|P(\sqrt{n}(\bar{X}_n - \mu) < x) - P^*\left(\sqrt{n}\left(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\right) < x\right)\|_\infty$$

$$= O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})\ a.s.$$

(B) *If $E|X_1|^3 < \infty$, $EN_n = n + O(\sqrt{n})$ and $DN_n = O(n)$, then*

$$\left\| P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\right) - P^*\left(\frac{\sqrt{n}}{s_n}\left(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\right) < x\right)\right\|_\infty = O(n^{-\frac{1}{2}})\ a.s.$$

and

$$(1 + |x|^3)\left|P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\right) - P^*\left(\frac{\sqrt{n}}{s_n}(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n) < x\right)\right|$$

$$= O\left(n^{-\frac{1}{2}}\right)\ a.s.$$

Further, the main result on the same convergence problem for the approximation $\sqrt{N_n}(\bar{X}^*_{N_n} - \bar{X}_n)$ or $\frac{\sqrt{N_n}}{s_n}(\bar{X}^*_{N_n} - \bar{X}_n)$ is the following theorem.

Theorem 2. *Let $N_n, X_1, X_2, \ldots$ be as in Theorem 1 and $E(N_n^{-\frac{1}{2}}) = O(n^{-\frac{1}{2}})$.*

(A) *If $EX_1^4 < \infty$, then*

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}} \| P(\sqrt{n}(\bar{X}_n - \mu) < x) - P^*(\sqrt{N_n}(\bar{X}^*_{N_n} - \bar{X}_n) < x) \|_\infty \le \frac{\sqrt{D((X_1-\mu)^2)}}{2\sigma^2\sqrt{\pi e}} \ a.s.$$

(B) *If $E|X_1|^3 < \infty$, then*

$$\left\| P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\right) - P^*\left(\frac{\sqrt{N_n}}{s_n}(\bar{X}^*_{N_n} - \bar{X}_n) < x\right)\right\|_\infty = O(n^{-\frac{1}{2}}) \ a.s.$$

and

$$(1+|x|^3)\left| P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\right) - P^*\left(\frac{\sqrt{N_n}}{s_n}(\bar{X}^*_{N_n} - \bar{X}_n) < x\right)\right| = O(n^{-\frac{1}{2}}) \ a.s.$$

For the proof of our theorems we will need some facts and easily derived results.

Lemma 1. *For every $c > 0$, we have*

$$\|x\phi(cx)\|_\infty = \frac{1}{c\sqrt{2\pi e}}$$

and

$$\|(1+|x|^3)x\phi(cx)\|_\infty \le \frac{1}{c\sqrt{2\pi e}} + \frac{2\sqrt{2}}{e^2c^4\sqrt{\pi}},$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Lemma 2. [6, Lemma 6.3.2, p. 186] *For every $c > 0$, we have*

$$|\Phi(x) - \Phi(cx)| \le \min\{1, |x|\phi(\min(1,c)x)|1-c|\},$$

where $\Phi(x)$ is the standard normal distribution function.

By the proof of Theorem 1 in [9], we have

Lemma 3. *With the notation and assumptions as in the previous section, we have: if $EX_1^4 < \infty$, then*

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}}|s_n^2 - \sigma^2| = \sqrt{2D((X_1-\mu)^2)} \ a.s.$$

and

$$\left\|\Phi\left(\frac{x}{s_n}\right) - \Phi\left(\frac{x}{\sigma}\right) - \left(\frac{1}{s_n} - \frac{1}{\sigma}\right)x\phi\left(\frac{x}{\sigma}\right)\right\|_\infty = O(n^{-1}\log\log n) \ a.s.$$

Lemma 4. [6, Lemma 6.3.1, p. 186] *Let $X_1, X_2, \ldots$ be i.i.d. random variables with $EX_1 = 0$. If $E|X_1^3| < \infty$, then*

$$(1+|x|^3)|P(S_n < x\sigma\sqrt{n}) - \Phi(x)| \leq \frac{c\rho}{\sigma^3\sqrt{n}},$$

where $S_n = \sum_{i=1}^n X_i$, $\rho = E|X_1 - \mu|^3$ and c is absolute constant, $c \leq 30.5378$.

Also, in the proofs, we shall use the following versions of Theorems 6.2.1 and 6.3.1 in [6].

Theorem A. *Let $N, X_1, X_2, \ldots$ be independent random variables, where N takes values among the natural numbers and $X_1, X_2, \ldots$ are identically distributed with $EX_1 = 0$. If $E|X_1^3| < \infty$, then for all $a \in (0,1)$, we have*

$$\left\| P(S_N < x) - \Phi\left(\frac{x}{\sigma\sqrt{EN}}\right)\right\|_\infty \leq \frac{K\rho}{\sigma^3\sqrt{a^3 EN}} + Q_1(a)E\left|\frac{N}{EN} - 1\right|,$$

where K is the universal appearing in the Berry–Esséen bound, $S_N = \sum_{i=1}^N X_i$ and $Q_1(a) = \max\left\{\frac{1}{1-a}, \frac{1}{\sqrt{2\pi e a}}\cdot\frac{1}{1+\sqrt{a}}\right\}$.

Theorem B. *With $N, X_1, X_2, \ldots$ as in Theorem A, if $E|X_1^3| < \infty$ and $EN^2 < \infty$, then for all $a \in (0,1)$, $b \in (1,\infty)$,*

$$(1+|x|^3)\left|P(S_N < x) - \Phi\left(\frac{x}{\sigma\sqrt{EN}}\right)\right| \leq K_1(a,3)\frac{\rho}{\sigma^3\sqrt{EN}}$$
$$+K_2(a,b,3)\max\left\{E\left|\frac{N}{EN} - 1\right|, \frac{(DN)^{\frac{3}{4}}}{(EN)^{\frac{3}{2}}}\right\},$$

where

$$K_1(a,3) = c + 0.7655a^{-\frac{3}{2}},\ c \leq 30.5378,$$

$$K_2(a,b,3) = \max\left\{\frac{w(b,3)}{a+\sqrt{a}}, \frac{v(3)}{1-a}\right\} + \frac{b^2u(3)}{(b-1)^2} + \frac{1}{1-a},$$

$$w(b,3) = \left\|(1+|x|^3)x\phi\left(\frac{x}{\sqrt{b}}\right)\right\|_\infty,$$

$$v(3) = \left\|(1+|x|^3)\min\left\{1, \frac{\phi(x)}{|x|}\right\}\right\|_\infty < 1.2936,$$

$$u(3) = \left\|(1+|x|^3)\min\left\{1, \sqrt{\frac{2}{\pi}}\frac{\Gamma(2)}{|x|^3}\right\}\right\|_\infty < 2.5958.$$

Remark. If N_n is a Poisson variable with $EN_n = n$, and $EX_1^4 < \infty$, then

$$\limsup_{n\to\infty} n^{\frac{1}{2}} (\log\log n)^{-\frac{1}{2}} \| P\left(\sqrt{n}\left(\bar{X}_n - \mu\right) < x\right) - P^*(\sqrt{n}(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n) < x)\|_\infty$$

$$\leq \frac{\sqrt{D((X_1 - \mu)^2)}}{2\sigma^2\sqrt{\pi e}} \text{ a.s.}$$

However, if $E|X_1^3| < \infty$, we only have

$$\limsup_{n\to\infty} n^{\frac{1}{2}} \left\| P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\right) - P^*\left(\frac{\sqrt{n}}{s_n}(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n) < x\right)\right\|_\infty$$

$$\leq \frac{K\rho}{\sigma^3}\left(1 + \frac{1}{\sqrt{a^3}}\right) + Q_1(a)\sqrt{\frac{2}{\pi}} \text{ a.s. } \forall a \in (0,1)$$

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(1 + |x|^3) \left| P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\right) - P^*\left(\frac{\sqrt{n}}{s_n}\left(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\right) < x\right)\right|$$

$$\leq \frac{\rho}{\sigma^3}(c + K_1(a,3)) + \sqrt{\frac{2}{\pi}} K_2(a,b,3) \text{ a.s. } \forall a \in (0,1),\ \forall b \in (1,\infty),$$

$$\limsup_{n\to\infty} n^{\frac{1}{2}} \left\| P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\right) - P^*\left(\frac{\sqrt{N_n}}{s_n}(\bar{X}_{N_n}^* - \bar{X}_n) < x\right)\right\|_\infty$$

$$\leq \frac{2K\rho}{\sigma^3} \text{ a.s.}$$

and

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(1 + |x|^3) \left| P\left(\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\right) - P^*\left(\frac{\sqrt{N_n}}{s_n}(\bar{X}_{N_n}^* - \bar{X}_n) < x\right)\right|$$

$$\leq \frac{2c\rho}{\sigma^3} \text{ a.s.}$$

Acknowledgement. The author would like to express many thanks to the referee for valuable comments and to Professor D. Q. Luu for helping him in writing the final version according to the referee's suggestion.

References

1. K. B. Athreya, Bootstrap of the mean in the infinite variance case, *Ann. Statist.* **15** (1987) 724–731.
2. S. Csörgö and D. M. Mason, Bootstrapping empirical functions, *Ann. Statist.* **17** (1989) 1447–1471.

3. E. Giné and J. Zinn, Necessary conditions for the bootstrap of the mean, *Ann. Statist.* **17** (1989) 684–691.
4. P. Hall, Asymptotic properties of the bootstrap for heavy-tailed distributions, *Ann. Probab.* **18**(3) (1990) 1342–1360.
5. K. B. Knight, On the bootstrap of the sample mean in the infinite variance case, *Ann. Statist.* **17** (1989) 1168–1175.
6. V. M. Kruglov and V. Yu. Korolev, *The Limit Theorem for Random Sums*, Moscow University, Moscow, 1990 (Russian).
7. E. Mammen, Bootstrap, wild bootstrap, and asymptotic normality, *Prob. Theory Relat. Fields* **93** (1992) 439–455.
8. Nguyen Van Toan, Wild bootstrap and asymptotic normality, *Scientific Information, College of Science, University of Hue* **10** (1997) 48–52.
9. K. Singh, On the asymptotic accuracy of Efron's bootstrap, *Ann. Statist.* **9**(6) (1981) 1187–1195.
10. Tran Manh Tuan and Nguyen Van Toan, On the asymptotic theory for the bootstrap with random sample size (preprint).
11. Tran Manh Tuan and Nguyen Van Toan, An asymptotic normality theorem of the bootstrap sample with random sample size, *Hanoi National Univ. J. Sc., Nat. Sci.* **XIV**(1) (1998) 1–7.