ACTA MATHEMATICA VIETNAMICA
Volume 25, Number 2, 2000, pp. 161–179

RATE OF CONVERGENCE IN BOOTSTRAP APPROXIMATIONS WITH RANDOM SAMPLE SIZE

NGUYEN VAN TOAN

Abstract. We give the convergence rates of the random sample size bootstrap approximations of the distribution of the sample mean. Using the quantity $\alpha_n^2 = E\Big[\Big(\frac{N_n}{n} - 1\Big)^2\Big]$, where N_n is the random sample size, we obtain a convergence rate $O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$ if $\alpha_n = O\Big(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}\Big)$.

1. Introduction

Efron [2] introduced a very general resampling procedure, called the bootstrap, for estimating the distributions of statistics based on independent observations. It may be explained briefly as follows: Let $X_1, \dots, X_n$ be independent identically distributed (i.i.d.) observations with distribution F; let $\theta = \theta(F)$ be a parameter of interest, and let $\theta_n = \theta_n(X_1, \dots, X_n)$ be an estimator of θ. The bootstrap principle is to estimate the unknown distribution of θ_n by $\hat{\theta}_n$ where $\hat{\theta}_n$ is distributed as $\theta_n(X^*_{n1}, \dots, X^*_{nn})$ and $X^*_{n1}, \dots, X^*_{nn}$ are i.i.d. from the empirical distribution of $(X_1, \dots, X_n)$.

In sufficiently regular cases, the bootstrap approximation to an unknown distribution function has been established as an improvement over the simpler normal approximation (see [1], [3-4]). In the case where the bootstrap sample size N is in itself a random variable, Mammen [7] has considered bootstrap with a Poisson random sample size which is independent of the sample. Our work on this problem is limited to [8-11] and [13-14]. In these references we find sufficient conditions for random sample size that random sample size bootstrap distribution can be used to approximate the sampling distribution. The purpose of this paper is to study the convergence rates to zero of the discrepancy between the actual

Received January 26, 1999; in revised form November 2, 1999.
1991 Mathematics Subject Classification. Primary: 62E20; Secondary: 62G05, 62G15.
Key words and phrases. Bootstrap, Berry-Esseen bound, Liapounov's inequality, random sample size.
This research is supported by the National Basic Research Program in Natural Sciences, Vietnam.

distribution of the sample mean and the random sample size bootstrap approximations of it.

We derive our results by studying the rates of convergence of the bootstrap approximations with a random sample size. Before giving more detail, it is necessary to introduce some notations. Let $X, X_1, X_2, \dots$ be i.i.d. random variables with finite variance. For any fixed natural number n, denote by F_n the empirical distribution of $(X_1, \dots, X_n)$ and by $X^*_{n1}, X^*_{n2}, \dots$ i.i.d. random variables with the distribution F_n. Let N_n be a positive integer-valued random variable independent of $X_1, X_2, \dots$. Set

$$\begin{aligned}
&\mu = EX \text{ is the expectation of } X,\\
&0 < \sigma^2 = DX \text{ is the variance of } X,\\
&\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i,\\
&s_n^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,\\
&\bar{X}_n^{*N_n} = n^{-1}\sum_{i=1}^{N_n} X^*_{ni},\\
&\bar{X}^*_{N_n} = N_n^{-1}\sum_{i=1}^{N_n} X^*_{ni}.
\end{aligned}$$

In the case where $N_n \to_p \infty$ as $n \to \infty$, where $\to_p$ denotes convergence in probability, the random sample size bootstrap distribution of $\sqrt{N_n}(\bar{X}^*_{N_n} - \bar{X}_n)$ and $\dfrac{\sqrt{N_n}}{s_n}(\bar{X}^*_{N_n} - \bar{X}_n)$ can be used to approximate the sampling distributions of $\sqrt{n}(\bar{X}_n - \mu)$ and $\dfrac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$, respectively (see [9], [13]). When $\dfrac{N_n}{n} \to_p 1$ as $n \to \infty$, the random sample size bootstrap distribution of $\sqrt{n}\big(\bar{X}_n^{*N_n} - \dfrac{N_n}{n}\bar{X}_n\big)$ and $\dfrac{\sqrt{n}}{s_n}\big(\bar{X}_n^{*N_n} - \dfrac{N_n}{n}\bar{X}_n\big)$ can be used as given in [8] and [14].

Singh [12] has studied the uniform convergence to zero of the discrepancy between the actual distribution of $\sqrt{n}(\bar{X}_n - \mu)$ and the bootstrap approximation of it. The same convergence problem for the distribution of $\dfrac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$ was also studied by Singh in [12]. In this paper we extend the results of Singh [12] to the random sample size bootstrap approximations.

Namely, we study the uniform and non-uniform convergence rates to zero of the discrepancy between the actual distribution of $\sqrt{n}(\bar{X}_n - \mu)$ and its random sample size bootstrap approximations. We also investigate the same convergence problem for the distribution of $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$.

2. Results

For simplicity we let

$$\Delta_1^{*N_n} = \sup_{-\infty<x<+\infty} \Big| P\big[\sqrt{n}(\bar{X}_n - \mu) < x\big] - P^*\Big[\sqrt{n}\Big(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\Big) < x\Big]\Big|,$$

$$\Delta_2^{*N_n} = \sup_{-\infty<x<+\infty} \Big| P\Big[\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\Big] - P^*\Big[\frac{\sqrt{n}}{s_n}\Big(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\Big) < x\Big]\Big|,$$

$$\delta_1^{*N_n}(x) = \Big(1 + \Big|\frac{x}{\sigma}\Big|^3\Big)\Big| P\big[\sqrt{n}(\bar{X}_n - \mu) < x\big] - P^*\Big[\sqrt{n}\Big(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\Big) < x\Big]\Big|,$$

$$\delta_2^{*N_n}(x) = \Big(1 + \Big|\frac{x}{\sigma}\Big|^3\Big)\Big| P\Big[\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\Big] - P^*\Big[\frac{\sqrt{n}}{s_n}\Big(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\Big) < x\Big]\Big|,$$

where P^* denotes conditional probability $P(\dots|X_1, \dots, X_n)$, and let

$$\alpha_n^2 = E\Big[\Big(\frac{N_n}{n} - 1\Big)^2\Big], \quad \beta_n = \max(n^{-\frac{1}{2}}, \alpha_n),$$

$$\gamma_n = \max(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}, \alpha_n).$$

Our main result are presented in the next two theorems. In Theorem 2.1, we study the convergence rates to zero of $\Delta_1^{*N_n}$, $\delta_1^{*N_n}(x)$, $\Delta_2^{*N_n}$ and $\delta_2^{*N_n}(x)$.

Theorem 2.1. *Suppose that the sequence N_n of positive integer-valued random variables satisfies the condition*

(a) $\alpha_n \to 0$ *as $n \to \infty$.*

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

$$\Delta_1^{*N_n} = O(\gamma_n) \ \text{a.s.}, \tag{2.1}$$

$$\delta_1^{*N_n}(x) = O(\gamma_n) \ \text{a.s.}. \tag{2.2}$$

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

$$\Delta_2^{*N_n} = O(\beta_n) \ \textit{a.s.}, \tag{2.3}$$

$$\delta_2^{*N_n}(x) = O(\beta_n) \ \textit{a.s.}. \tag{2.4}$$

In particular, we have the following theorem which is analogous to Theorem 2.1 of [10].

Theorem 2.1'.
A) *If* $EX^4 < \infty$ *and if the sequence* N_n *of positive integer-valued random variables satisfies the condition*
(b) $\alpha_n = O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}})$*, then*

$$\Delta_1^{*N_n} = O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) \ \textit{a.s.}, \tag{2.1'}$$

$$\delta_1^{*N_n}(x) = O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) \ \textit{a.s.}. \tag{2.2'}$$

B) *If* $E|X^3| < \infty$ *and if the sequence* N_n *of positive integer-valued random variables satisfies the condition*
(c) $\alpha_n = O(n^{-\frac{1}{2}})$*, then*

$$\Delta_2^{*N_n} = O(n^{-\frac{1}{2}}) \ \textit{a.s.}, \tag{2.3'}$$

$$\delta_2^{*N_n}(x) = O(n^{-\frac{1}{2}}) \ \textit{a.s.}. \tag{2.4'}$$

Let

$$R_{N_n}^{*1} = \sup_{-\infty<x<+\infty} \Big| P\big[\sqrt{n}(\bar{X}_n - \mu) < x\big] - P^*\Big[\sqrt{N_n}(\bar{X}^*_{N_n} - \bar{X}_n) < x\Big]\Big|,$$

$$R_{N_n}^{*2} = \sup_{-\infty<x<+\infty} \Big| P\Big[\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\Big] - P^*\Big[\frac{\sqrt{N_n}}{s_n}(\bar{X}^*_{N_n} - \bar{X}_n) < x\Big]\Big|,$$

$$r_{N_n}^{*1}(x) = \big(1 + |x|^3\big)\Big| P\big[\sqrt{n}(\bar{X}_n - \mu) < x\big] - P^*\big[\sqrt{N_n}(\bar{X}^*_{N_n} - \bar{X}_n) < x\big]\Big|,$$

$$r_{N_n}^{*2}(x) = \big(1 + |x|^3\big)\Big| P\Big[\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) < x\Big] - P^*\Big[\frac{\sqrt{N_n}}{s_n}\Big(\bar{X}^*_{N_n} - \frac{N_n}{n}\bar{X}_n\Big) < x\Big]\Big|,$$

and

$$\eta_n = E[N_n^{-\frac{1}{2}}], \ \zeta_n = \max(n^{-\frac{1}{2}}, \eta_n), \ \theta_n = \max(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}, \eta_n).$$

The result on the same convergence problem for $R^{*1}_{N_n}$, $r^{*1}_{N_n}(x)$, $R^{*2}_{N_n}$ and $r^{*2}_{N_n}(x)$ is the following theorem.

Theorem 2.2.
A) *If* $EX^4 < \infty$, *then*

$$R^{*1}_{N_n} = O(\theta_n) \ \ a.s., \tag{2.5}$$

$$r^{*1}_{N_n}(x) = O(\theta_n) \ \ a.s.. \tag{2.6}$$

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

$$R^{*2}_{N_n} = O(\zeta_n) \ \ a.s., \tag{2.7}$$

$$r^{*2}_{N_n}(x) = O(\zeta_n) \ \ a.s. \tag{2.8}$$

In particular, we have the following theorem which is analogous to Theorem 2.2 of [10].

Theorem 2.2'. *Suppose that* $E[N_n^{-1}] = O(n^{-1})$.
A) *If* $EX^4 < \infty$, *then*

$$\limsup_{n\to\infty} n^{\frac{1}{2}} (\log\log n)^{-\frac{1}{2}} R^{*1}_{N_n} \le \frac{\sqrt{D[(X-\mu)^2]}}{2\sigma^2\sqrt{\pi e}}, \tag{2.5'}$$

$$r^{*1}_{N_n}(x) = O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) \ \ a.s.. \tag{2.6'}$$

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

$$R^{*2}_{N_n} = O(n^{-\frac{1}{2}}) \ \ a.s., \tag{2.7'}$$

$$r^{*2}_{N_n}(x) = O(n^{-\frac{1}{2}}) \ \ a.s.. \tag{2.8'}$$

3. Proofs

In this section we give the proofs of the above theorems. Towards the end some remarks concerning these conclusions will be given.

For the simplicity of the proofs we state some facts and easily derived results will be given.

Lemma 3.1. *For every $c > 0$ we have*

$$\sup_{-\infty<x<+\infty} |x\phi(cx)| = \frac{1}{c\sqrt{2\pi e}}\,, \tag{3.1}$$

$$\sup_{-\infty<x<+\infty} |(1+|x|^3)x\phi(cx)| \leq \frac{1}{c\sqrt{2\pi e}} + \frac{2\sqrt{2}}{e^2c^4\sqrt{\pi}}\,, \tag{3.2}$$

$$|\Phi(x) - \Phi(cx)| \leq \min\{1, |x|\phi(\min(1,c)x)|1-c|\}, \tag{3.3}$$

where $\Phi(x)$ and $\phi(x)$ are the standard normal distribution function and density, respectively.

By the proof of Theorem 1 of [12] we have

Lemma 3.2. *Let $X, X_1, X_2, \ldots$ be i.i.d. random variables. If $EX^4 < \infty$, then*

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}}|s_n^2 - \sigma^2| = \sqrt{2D[(X-\mu)^2]} \text{ a.s.}, \tag{3.4}$$

$$\sup_{-\infty<x<+\infty} \Big|\Phi\Big(\frac{x}{s_n}\Big) - \Phi\Big(\frac{x}{\sigma}\Big) - \Big(\frac{1}{s_n} - \frac{1}{\sigma}\Big)x\phi\Big(\frac{x}{\sigma}\Big)\Big| = O(n^{-1}\log\log n) \text{ a.s.} \tag{3.5}$$

We will also need the following results of Kruglov and Korolev [6].

Lemma 3.3 [6, Lemma 6.3.1]. *Let $X, X_1, X_2, \ldots$ be i.i.d. random variables with $EX = 0$. If $E|X^3| < \infty$, then*

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

where $S_n = \sum\limits_{i=1}^{n} X_i$, $\rho = E|X^3|$ and c is absolute constant, $c \leq 30.5378$.

Theorem 3.1 [6, Theorem 6.2.1]. *Let $N, X, X_1, X_2, \ldots$ be independent random variables, where N takes values among the natural numbers and $X, X_1, X_2, \ldots$ are identically distributed with $EX = 0$. If $E|X^3| < \infty$, then for all $a \in (0,1)$*

$$\sup_{-\infty<x<+\infty}\Big|P(S_N<x)-\Phi\Big(\frac{x}{\sigma\sqrt{EN}}\Big)\Big|\le\frac{K\rho}{\sigma^3\sqrt{a^3EN}}+Q(a)E\Big|\frac{N}{EN}-1\Big|,$$

where K *is the universal appearing in the Berry-Esseen bound,* $S_N=\sum_{i=1}^{N}X_i$ *and* $Q(a)=\max\Big\{\frac{1}{1-a},\frac{1}{\sqrt{2\pi ea}},\frac{1}{1+\sqrt{a}}\Big\}$.

Theorem 3.2 [6, Theorem 6.3.1]. *With* $N,X,X_1,X_2,\dots$ *as in Theorem 3.1, if* $E|X^3|<\infty$ *and* $EN^2<\infty$*, then for all* $a\in(0,1)$, $b\in(1,\infty)$

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

where

$$\begin{aligned}
K_1(a,3)&=c+0.7655a^{-\frac{3}{2}},\quad c\le 30.5378,\\
K_2(a,b,3)&=\max\Big\{\frac{w(b,3)}{a+\sqrt{a}}\ ,\ \frac{v(3)}{1-a}\Big\}+\frac{b^2u(3)}{(b-1)^2}+\frac{1}{1-a},\\
w(b,3)&=\sup_{-\infty<x<+\infty}\Big|(1+|x|^3)x\phi\Big(\frac{x}{\sqrt{b}}\Big)\Big|,\\
v(3)&=\sup_{-\infty<x<+\infty}\Big|(1+|x|^3)\min\Big\{1,\frac{\phi(x)}{|x|}\Big\}\Big|<1.2936,\\
u(3)&=\sup_{-\infty<x<+\infty}\Big|(1+|x|^3)\min\Big\{1,\sqrt{\frac{2}{\pi}}\frac{\Gamma(2)}{|x|^3}\Big\}\Big|<2.5958.
\end{aligned}$$

Proof of Theorem 2.1. We first note that

$$EN_n\sim n \text{ as } n\to\infty, \tag{3.6}$$

by the inequality

$$|EN_n-n|\le E|N_n-n|\le\sqrt{E[(N_n-n)^2]}\ \text{(by Liapounov's inequality)}, \tag{3.7}$$

and condition (a).

Part A. It is worth noting that

$$\Delta_1^{*N_n} \le A_1 + A_2 + A_3 + A_4 + A_5,$$

where

$$A_1 = \sup_{-\infty<x<+\infty}\Big|P\big[\sqrt{n}(\bar{X}_n - \mu) < x\big] - \Phi\Big(\frac{x}{\sigma}\Big)\Big|,$$
$$A_2 = \sup_{-\infty<x<+\infty}\Big|\Phi\Big(t_n\frac{x}{\sigma}\Big) - \Phi\Big(\frac{x}{\sigma}\Big)\Big|,$$
$$A_3 = \sup_{-\infty<x<+\infty}\Big|\Phi\Big(t_n\frac{x}{s_n}\Big) - \Phi\Big(t_n\frac{x}{\sigma}\Big) - t_n\Big(\frac{1}{s_n} - \frac{1}{\sigma}\Big)x\phi\Big(t_n\frac{x}{\sigma}\Big)\Big|,$$
$$A_4 = \sup_{-\infty<x<+\infty}\Big|t_n\Big(\frac{1}{s_n} - \frac{1}{\sigma}\Big)x\phi\Big(t_n\frac{x}{\sigma}\Big)\Big|,$$
$$A_5 = \sup_{-\infty<x<+\infty}\Big|P^*\Big[\sqrt{n}\Big(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\Big) < x\Big] - \Phi\Big(t_n\frac{x}{s_n}\Big)\Big|$$

and

$$t_n = \sqrt{\frac{n}{EN_n}}.$$

Firstly, due to the Berry-Esseen bound, we have

$$A_1 \le \frac{K\rho}{\sigma^3\sqrt{n}}, \tag{3.8}$$

where K is defined as in Theorem 3.1. By Theorem 3.1,

$$A_5 \le \frac{K\hat{\rho}_n}{s_n^3\sqrt{a^3EN_n}} + Q(a)E\Big|\frac{N_n}{EN_n} - 1\Big|, \tag{3.9}$$

where $\hat{\rho}_n = \dfrac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_n|^3$.

The condition $EX^4 < \infty$ implies $s_n^2 \to \sigma^2$ and $\hat{\rho}_n \to \rho$ a.s., and hence

$$\frac{K\hat{\rho}_n}{s_n^3\sqrt{a^3EN_n}} = O(n^{-\frac{1}{2}}) \text{ a.s.,} \tag{3.10}$$

because of (3.6). Further, by (3.1), (3.6) and Lemma 3.2 we see that under the condition $EX^4 < \infty$ we get

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}}A_4 = \frac{\sqrt{D[(X-\mu)^2]}}{2\sigma^2\sqrt{\pi e}} \text{ a.s.,} \tag{3.11}$$

$$A_3 = O(n^{-1}\log\log n) \text{ a.s..} \tag{3.12}$$

Now, by Lemma 3.1 we see that

$$A_2 \le \frac{|t_n - 1|}{\sqrt{2\pi e}\min(1, t_n)}. \tag{3.13}$$

But we have

$$|t_n - 1| \le |t_n^2 - 1| = \frac{|n - EN_n|}{EN_n}. \tag{3.14}$$

Therefore, (2.1) follows from (3.6)-(3.14) and the inequality

$$E|N_n - EN_n| \le 2E|N_n - n|. \tag{3.15}$$

To show (2.2) we note that

$$\delta_1^{*N_n}(x) \le B_1(x) + B_2(x) + B_3(x) + B_4(x),$$

where

$$\begin{aligned}
B_1(x) &= \Big(1 + \Big|\frac{x}{\sigma}\Big|^3\Big)\Big|P\big[\sqrt{n}(X_n - \mu) < x\big] - \Phi\Big(\frac{x}{\sigma}\Big)\Big|,\\
B_2(x) &= \Big(1 + \Big|\frac{x}{\sigma}\Big|^3\Big)\Big|\Phi\Big(t_n\frac{x}{\sigma}\Big) - \Phi\Big(\frac{x}{\sigma}\Big)\Big|,\\
B_3(x) &= \Big(1 + \Big|\frac{x}{\sigma}\Big|^3\Big)\Big|\Phi\Big(t_n\frac{x}{s_n}\Big) - \Phi\Big(\frac{x}{\sigma}\Big)\Big|,\\
B_4(x) &= \Big(1 + \Big|\frac{x}{\sigma}\Big|^3\Big)\Big|P^*\Big[\sqrt{n}\Big(\bar{X}_n^{*N_n} - \frac{N_n}{n}\bar{X}_n\Big) < x\Big] - \Phi\Big(t_n\frac{x}{s_n}\Big)\Big|.
\end{aligned}$$

Since $EX^4 < \infty$, Lemma 3.3 implies

$$B_1(x) \le \frac{c\rho}{\sigma^3\sqrt{n}}. \tag{3.16}$$

We now apply Lemma 3.1 to get

$$\begin{aligned}
B_2(x) &\le \Big(1 + \Big|\frac{x}{\sigma}\Big|^3\Big)\min\Big\{1, \Big|\frac{x}{\sigma}\Big|\phi\Big(\min(1, t_n)\frac{x}{\sigma}\Big)|1 - t_n|\Big\}\\
&\le \Big[\frac{1}{\sqrt{2\pi e}\min(1, t_n)} + \frac{2\sqrt{2}}{e^2\sqrt{\pi}[\min(1, t_n)]^4}\Big]|1 - t_n|
\end{aligned} \tag{3.17}$$

and

$$(3.18)\quad B_3(x) \le \Big(1+\Big|\frac{x}{\sigma}\Big|^3\Big)\min\Big\{1, \Big|t_n\frac{x}{\sigma}\Big|\phi\Big(\min\big(1,\frac{\sigma}{s_n}\big)t_n\frac{x}{\sigma}\Big)\Big|1-\frac{\sigma}{s_n}\Big|\Big\}$$

$$\le \frac{1+|\frac{x}{\sigma}|^3}{1+|t_n\frac{x}{\sigma}|^3}\left[\frac{1}{\sqrt{2\pi e}\min\big(1,\frac{\sigma}{s_n}\big)}+\frac{2\sqrt{2}}{e^2\sqrt{\pi}\Big[\min\big(1,\frac{\sigma}{s_n}\big)\Big]^4}\right]\Big|1-\frac{\sigma}{s_n}\Big|.$$

By (3.7) and (3.14) we get

$$(3.19)\qquad B_2(x) = O(\alpha_n).$$

Since $s_n^2 \to \sigma^2$ a.s. and because of (3.6), a repeated application of Lemma 3.2 enables us to write

$$(3.20)\qquad B_3(x) = O(n^{-\frac{1}{2}}(\log\log n)^{\frac{1}{2}}) \text{ a.s..}$$

Theorem 3.2 now shows that

(3.21)

$$B_4(x) \le \frac{1+\Big|\frac{x}{\sigma}\Big|^3}{1+\Big|\frac{x}{s_n}\Big|^3}\Big\{K_1(a,3)\frac{\hat{\rho}_n}{s_n^3\sqrt{EN_n}}$$
$$+K_2(a,b,3)\max\Big\{E\Big|\frac{N_n}{EN_n}-1\Big|, \frac{(DN_n)^{\frac{3}{4}}}{(EN)^{\frac{3}{2}}}\Big\}\Big\}$$

for all $a \in (0,1)$, $b \in (1,\infty)$. From (3.7), (3.10), (3.15) and the inequality

$$(3.22)\qquad DN_n \le E\left[(N_n-n)^2\right],$$

it follows that

$$(3.23)\qquad B_4(x) = O(\beta_n) \text{ a.s..}$$

Combining (3.16),(3.19),(3.20) and (3.23) we obtain (2.2).

Part B. In order to prove (2.3) we note that

$$\Delta_2^{*N_n} \le A_1 + A_2 + A_5,$$

where $A_1,\ A_2$ and A_5 are defined as in Part A, and if $EX^4 < \infty$ is replaced by $E|X^3| < \infty$. Then (3.8)-(3.10), (3.13)-(3.15) also hold. Hence (2.3) follows from (3.7).

To derive (2.4) note that

$$\delta_2^{*N_n}(x) \le C_1(x) + C_2(x) + C_3(x),$$

where

$$\begin{aligned}
C_1(x) &= (1+|x|^3)\Big|P\Big(\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} < x\Big) - \Phi(x)\Big|,\\
C_2(x) &= (1+|x|^3)\,|\Phi(t_n x) - \Phi(x)|\,,\\
C_3(x) &= (1+|x|^3)\Big|P^*\Big[\frac{\sqrt{n}\left(\bar{X}_n^{*N_n} - \dfrac{N_n}{n}\bar{X}_n\right)}{s_n} < x\Big] - \Phi\left(t_n x\right)\Big|.
\end{aligned}$$

Since $E|X^3| < \infty$, by Lemma 3.3 we immediately have

$$C_1(x) \le \frac{c\rho}{\sigma^3\sqrt{n}},$$

and by applying Theorem 3.2,

$$C_3(x) \le K_1(a,3)\frac{\hat{\rho}_n}{s_n^3\sqrt{EN_n}} + K_2(a,b,3)\max\Big\{E\Big|\frac{N_n}{EN_n}-1\Big|, \frac{(DN_n)^{\frac{3}{4}}}{(EN_n)^{\frac{3}{2}}}\Big\}$$

for all $a \in (0,1)$, $b \in (1,\infty)$. From (3.7), (3.10), (3.15) and (3.22), we deduce that

$$C_3(x) = O(\beta_n) \text{ a.s..}$$

Finally, applying Lemma 3.1, $C_2(x)$ is bounded above by the right-hand side of (3.17), and hence, by (3.7), it follows that

$$C_2(x) = O(\alpha_n).$$

This with the previous estimates prove (2.4). The proof of Theorem 2.1 is now complete.

Proof of Theorem 2.1'.
Part A. If condition (b) holds, then

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

Therefore, (2.1') and (2.2') immediately follow from (2.1) and (2.2), respectively.

Part B. In the case where condition (c) holds, we have

$$\beta_n = O(n^{-\frac{1}{2}}),$$

and hence, by (2.3) and (2.4), we get (2.3') and (2.4'), respectively.

Proof of Theorem 2.2.

Part A. It is easy to check that

$$R^{*1}_{N_n} \leq A_1 + A_3 + A_4 + A'_5,$$

where $A_1,\ A_3$ and A_4 are defined as in the proof of Theorem 2.1, and

$$A'_5 = \sup_{-\infty<x<+\infty} \Big| P^*\big[\sqrt{N_n}\,\big(\bar{X}^*_{N_n} - \bar{X}_n\big) < x\big] - \Phi\Big(\frac{x}{s_n}\Big)\Big|.$$

Because of $EX^4 < \infty$ one gets (3.8), (3.11) and (3.12). By the law of total probability and the fact that N_n and $X_1, X_2, \ldots$ are independent, we get

$$\begin{aligned} A'_5 &\leq \sum_{m=1}^{\infty} P[N_n = m] \sup_{-\infty<x<+\infty} \Big| P^*\big[\sqrt{m}\,\big(\bar{X}^*_m - \bar{X}_n\big) < x\big] - \Phi\Big(\frac{x}{s_n}\Big)\Big| \\ &\leq \sum_{m=1}^{\infty} P[N_n = m] \frac{K\hat{\rho}_n}{s_n^3\sqrt{m}} \qquad \text{(by the Berry-Esseen bound)} \\ &= \frac{K\hat{\rho}_n}{s_n^3} E[N_n^{-\frac{1}{2}}]. \end{aligned}$$

Note that if $EX^4 < \infty,\ s_n^2 \to \sigma^2$ and $\hat{\rho}_n \to \rho$ a.s., and hence

$$A'_5 = O(\eta_n) \text{ a.s.}. \tag{3.24}$$

Combining theses results we obtain (2.5).

Using the argument leading to (2.2) and (3.24) and noting that

$$r^{*1}_{N_n}(x) \leq B_1(x) + \frac{1 + \left|\dfrac{x}{\sigma}\right|^3}{1 + \left|t_n \dfrac{x}{\sigma}\right|^3} B_3\Big(\frac{x}{t_n}\Big) + B'_4(x),$$

where $B_1(x)$ and $B_3(x)$ are defined as in the proof of Theorem 2.1, and

$$B_4'(x) = \Big(1 + \Big|\frac{x}{\sigma}\Big|^3\Big)\Big|P^*\big[\sqrt{N_n}\,(\bar{X}^*_{N_n} - \bar{X}_n) < x\big] - \Phi\Big(\frac{x}{s_n}\Big)\Big|,$$

we obtain (2.6).

Part B. If $E|X^3| < \infty$, $s_n^2 \to \sigma^2$ and $\hat{\rho}_n \to \rho$ a.s. Therefore, (2.7) follows from the inequality

$$R^{*2}_{N_n} \leq A_1 + A_5',$$

and (3.8) and (3.24).

To obtain (2.8) notice that

$$r^{*2}_{N_n}(x) \leq B_1(\sigma x) + \frac{1 + |x|^3}{1 + \Big|\dfrac{s_n}{\sigma}x\Big|^3} B_4'(s_n x),$$

and arguing as in the proof of result (2.6). This completes the proof of Theorem 2.2.

Proof of Theorem 2.2'. If $E[N_n^{-1}] = O(n^{-1})$, then by Liapounov' inequality we have $E[N_n^{-\frac{1}{2}}] = O(n^{-\frac{1}{2}})$. Using the proof of Theorem 2.2 we obtain Theorem 2.2'.

Remark 3.1. Theorem 2.1 can be stated more precisely as follows: Suppose that the sequence N_n of positive integer-valued random variables satisfies the condition:

(a) $\alpha_n \to 0$ as $n \to \infty$.

A) If $EX^4 < \infty$, then for all $a \in (0,1)$, $b \in (1,\infty)$

$$\limsup_{n\to\infty} \gamma_n^{-1}\Delta_1^{*N_n} \leq \frac{\sqrt{D[(X-\mu)^2]}}{2\sigma^2\sqrt{\pi e}} + \frac{1}{\sqrt{8\pi e}} + 2Q(a) \text{ a.s.},$$

$$\limsup_{n\to\infty} \gamma_n^{-1}\delta_1^{*N_n}(x) \leq \frac{e^{\frac{3}{2}}+4}{2e^2\sqrt{\pi}\sigma^2}\Big(\sqrt{D[(X-\mu)^2]} + \frac{\sqrt{2}}{2}\sigma^2\Big) + 2K_2(a,b,3) \text{ a.s.}.$$

B) If $E|X^3| < \infty$, then for all $a \in (0,1)$, $b \in (1,\infty)$

$$\limsup_{n\to\infty} \beta_n^{-1}\Delta_2^{*N_n} \leq \frac{(a^{\frac{3}{2}}+1)K\rho}{a^{\frac{3}{2}}\sigma^3} + \frac{1}{\sqrt{8\pi e}} + 2Q(a) \text{ a.s.},$$

$$\limsup_{n\to\infty} \beta_n^{-1}\delta_2^{*N_n}(x) \leq \big(c + K_1(a,3)\big)\frac{\rho}{\sigma^3} + \frac{e^{\frac{3}{2}}+4}{e^2\sqrt{8\pi}} + 2K_2(a,b,3) \text{ a.s.}.$$

Remark 3.2. Theorem 2.2 can be stated more precisely as follows:
A) If $EX^4 < \infty$, then

$$\limsup_{n\to\infty} \theta_n^{-1} R^{*1}_{N_n} \leq \frac{\sqrt{D[(X-\mu)^2]}}{2\sigma^2\sqrt{\pi e}} + \frac{K\rho}{\sigma^3} \text{ a.s.,}$$
$$\limsup_{n\to\infty} \theta_n^{-1} r^{*1}_{N_n}(x) \leq \frac{c\rho}{\sigma^3} + \frac{(e^{\frac{3}{2}}+4)\sqrt{D[(X-\mu)^2]}}{2\sigma^2 e^2\sqrt{\pi}} \text{ a.s..}$$

B) If $E|X^3| < \infty$, then

$$\limsup_{n\to\infty} \zeta_n^{-1} R^{*2}_{N_n} \leq \frac{2K\rho}{\sigma^3} \text{ a.s.,}$$
$$\limsup_{n\to\infty} \zeta_n^{-1} r^{*2}_{N_n}(x) \leq \frac{2c\rho}{\sigma^3} \text{ a.s..}$$

Remark 3.3. We can describe a general model that allows formulae such as the expansion of Edgeworth type. Let $X, X_1, X_2, \dots$ be independent and identically distributed random variables with mean μ, and put $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. We assume that the parameter of interest has form $\theta_0 = g(\mu)$, where $g : I\!R \to I\!R$ is a smooth function. The parameter estimate is $\hat{\theta} = g(\bar{X}_n)$, which is assumed to have asymptotic variance $\frac{\sigma}{n} = \frac{h(\mu)^2}{n}$, where h is a known, smooth function. It is supposed that $\sigma^2 \neq 0$. The sample estimate of σ^2 is $\hat{\sigma}^2 = h(\bar{X}_n)^2$.

We wish to estimate the distribution of either $\frac{\hat{\theta}-\theta_0}{\sigma}$ or $\frac{\hat{\theta}-\theta_0}{\hat{\sigma}}\cdot$ Note that both these statistics may be written in the form $A(\bar{X}_n)$, where

$$A(x) = \frac{g(x)-g(\mu)}{h(\mu)} \tag{3.25}$$

in the former case and

$$A(x) = \frac{g(x)-g(\mu)}{h(x)} \tag{3.26}$$

in the later. To construct the bootstrap distribution estimates, let $X^*_{n1}, \dots,$ X^*_{nm} denote a sample drawn randomly, with replacement, from $X_1, \dots,$

X_n, and write $\bar{X}^*_{nm} = \dfrac{1}{m}\sum_{i=1}^{m} X_{ni}*$ for the resample mean. Define

$$\hat{\theta}^*_{nm} = g(\bar{X}^*_{nm}),$$
$$\hat{\sigma}^*_{nm} = h(\bar{X}^*_{nm}).$$

Both $\dfrac{\hat{\theta}^*_{nm} - \hat{\theta}}{\hat{\sigma}}$ and $\dfrac{\hat{\theta}^*_{nm} - \hat{\theta}}{\hat{\sigma}^*}$ may be expressed as $\hat{A}(\bar{X}^*_{nm})$, where

$$\hat{A}(x) = \frac{g(x) - g(\bar{X}_n)}{h(\bar{X}_n)}, \tag{3.27}$$

$$\hat{A}(x) = \frac{g(x) - g(\bar{X}_n)}{h(x)} \tag{3.28}$$

in the respective cases. The bootstrap estimate of the distribution of $A(\bar{X}_n)$ is given by the distribution of $\hat{A}(\bar{X}^*_{nm})$, conditional on $X_1, \dots, X_n$.

Let N_n be a positive integer-valued random variable independent of $X_1, X_2, \dots$. The random sample size bootstrap estimate of the distribution of $A(\bar{X}_n)$ is given by the distribution of $\hat{A}(\bar{X}^*_{nN_n})$, conditional on $X_1, \dots, X_n$.

Theorem 2.2 of Hall [5] gave an Edgeworth expansion of the distribution of $A(\bar{X}_n)$, which may be stated as follows. Assume that for an integer $\nu \geq 1$, the function A has $\nu + 2$ continuous derivatives in a neighbourhood of μ, that $A(\mu) = 0$, that $E|X|^{\nu+2} < \infty$, that the characteristic function φ of X satisfies

$$\limsup_{|t|\to\infty} |\varphi(t)| < 1, \tag{3.29}$$

and that the asymptotic variance of $\sqrt{n}A(\bar{X}_n)$ equals 1. Then

$$P\{\sqrt{n}A(\bar{X}_n) \leq x\} = \Phi(x) + \sum_{j=1}^{\nu} n^{-\frac{j}{2}} \pi_j(x)\phi(x) + o(n^{-\frac{\nu}{2}})$$

uniformly in x, where π_j is a polynomial of degree of $3j - 1$, odd for even j and even for odd j, with coefficients depending on moments of X up to order $j + 2$. The form of π_j depends very much on the form of A. We denote π_j by p_j if A is given by (3.25), and by q_j if A is given by (3.26).

The bootstrap version of this result is described in Theorem 5.1 of Hall [5]. Let $\hat{A}$ be defined by either (27) or (28), put $\pi_j = p_j$ in the former case and $\pi_j = q_j$ in the latter, and let $\hat{\pi}_j$ be the polynomial obtained from π_j on replacing population moments by the corresponding sample moments. Arguing as in the proof of Theorem 5.1 [5], we can show the following result:

Let $\lambda > 0$ be given, and let $l = l(\lambda)$ denote a sufficiently large positive number (whose value we do not specify). Assume that g and h each have $\nu + 3$ bounded derivatives in a neighbourhood of μ, that $E|X|^l < \infty$, and that Cramer's condition (3.29) *holds. Then there exists a constant $C > 0$ such that*

(3.30)
$$P\Big[\Big\|P^*\{\sqrt{m}\hat{A}(\bar{X}^*_{nm}) \le x\} - \Phi(x) - \sum_{j=1}^{\nu} m^{-\frac{j}{2}}\hat{\pi}_j(x)\phi(x)\Big\| > Cm^{-\frac{\nu+1}{2}}\Big]$$
$$= O(n^{-\lambda}),$$
$$P\Big\{\max_{1\le j\le\nu}\sup_{-\infty<x<\infty}(1+|x|)^{-(3j-1)}|\hat{\pi}_j(x)| > C\Big\} = O(n^{-\lambda}).$$

If we take $\lambda > 1$ in this result and apply the Borel-Cantelli lemma, we may deduce from (3.30) that

$$\Big\|P^*\{\sqrt{m}\hat{A}(\bar{X}^*_{nm}) \le x\} - \Phi(x) - \sum_{j=1}^{\nu} m^{-\frac{j}{2}}\hat{\pi}_j(x)\phi(x)\Big\| = O(m^{-\frac{\nu+1}{2}}) \tag{3.31}$$

with probability one. And from (3.31) we deduce that

$$P^*\big\{\sqrt{N_n}\hat{A}(\bar{X}^*_{nN_n}) \le x\big\} = \Phi(x) + \sum_{j=1}^{\nu} E\big[N_n^{-\frac{j}{2}}\big]\hat{\pi}_j(x)\phi(x) + O\Big(E\big[N_n^{-\frac{\nu+1}{2}}\big]\Big)$$

with probability one.

Remark 3.4. If N_n is a Poisson variable with $EN_n = n$, then

$$DN_n = n,\ E|N_n - EN_n| = 2e^{-n}\frac{n^{n+1}}{n!} \sim \sqrt{\frac{2n}{\pi}},\ E\Big|\frac{N_n}{EN_n} - 1\Big| \sim \sqrt{\frac{2}{\pi n}},$$
$$\Big|\Phi\Big(t_n\frac{x}{\sigma}\Big) - \Phi\Big(\frac{x}{\sigma}\Big)\Big| = 0.$$

Therefore, if $EX^4 < \infty$, applying results (3.8), (3.9), (3.11)-(3.14), we obtain

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}}\Delta_1^{*N_n} \le \frac{\sqrt{D[(X-\mu)^2]}}{2\sigma^2\sqrt{\pi e}} \text{ a.s.,}$$

and from (3.16)-(3.18) and (3.21) we have

$$\limsup_{n\to\infty} n^{\frac{1}{2}}\big(\log\log n\big)^{-\frac{1}{2}}\delta_1^{*N_n}(x) \le \Big[\frac{1}{\sqrt{2\pi e}} + \frac{2\sqrt{2}}{e^2\sqrt{\pi}}\Big]\frac{\sqrt{D[(X-\mu)^2]}}{2\sigma^2} \text{ a.s.}$$

However, if $E|X^3| < \infty$, by the proof of Part B of Theorem 2.1 we only have

$$\limsup_{n\to\infty} n^{\frac{1}{2}}\Delta_2^{*N_n} \le \frac{K\rho}{\sigma^3}\Big(1+\frac{1}{\sqrt{a^3}}\Big) + Q(a)\sqrt{\frac{2}{\pi}} \text{ a.s.} \quad \forall a \in (0,1),$$

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(1+|x|^3)\delta_2^{*N_n}(x) \le \frac{\rho}{\sigma^3}(c+K_1(a,3)) + K_2(a,b,3)\sqrt{\frac{2}{\pi}} \text{ a.s.}$$

for all $a \in (0,1),\ b \in (1,\infty)$.

Remark 3.5. Also, in the case where N_n is a Poisson random variable with $EN_n = n$, we have

$$E[N_n^{-1}] = \sum_{m=1}^{\infty} e^{-n}\frac{n^m}{(m+1)!} = \frac{1-e^{-n}}{n},$$

$$E[N_n^{-\frac{1}{2}}] \le \sqrt{E[N_n^{-1}]} = \sqrt{\frac{1-e^{-n}}{n}}.$$

(where $N_n^{-1} = 0$ if $N_n = 0$). Hence, from the proof of Theorem 2.2 it follows that

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

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}}R_{N_n}^{*1} \le \frac{\sqrt{D[(X_1-\mu)^2]}}{2\sigma^2\sqrt{\pi e}} \text{ a.s.,}$$

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\log\log n)^{-\frac{1}{2}}r_{N_n}^{*1}(x) \le \Big[\frac{1}{\sqrt{2\pi e}} + \frac{2\sqrt{2}}{e^2\sqrt{\pi}}\Big]\frac{\sqrt{D[(X-\mu)^2]}}{2\sigma^2} \text{ a.s..}$$

B) If $E|X^3| < \infty$, then

$$\limsup_{n\to\infty} n^{\frac{1}{2}} R^{*2}_{N_n} \le \frac{2K\rho}{\sigma^3} \text{ a.s.},$$
$$\limsup_{n\to\infty} n^{\frac{1}{2}} r^{*2}_{N_n}(x) \le \frac{2c\rho}{\sigma^3} \text{ a.s..}$$

Remark 3.6. If the positive integer-valued random variable N_n has the uniform distribution function on $[1, n]$, then

$$E[N_n^{-1}] = \frac{1}{n}\sum_{m=1}^{n}\frac{1}{m} \sim \frac{\ln n}{n},$$
$$E[N_n^{-\frac{1}{2}}] \le \sqrt{E[N_n^{-1}]} \sim n^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}.$$

Hence, from the proof of Theorem 2.2 we get

A) If $EX^4 < \infty$, then

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\ln n)^{-\frac{1}{2}} R^{*1}_{N_n} \le \frac{K\rho}{\sigma^3} \text{ a.s.},$$
$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\ln n)^{-\frac{1}{2}} r^{*1}_{N_n}(x) \le \frac{c\rho}{\sigma^3} \text{ a.s..}$$

B) If $E|X^3| < \infty$, then

$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\ln n)^{-\frac{1}{2}} R^{*2}_{N_n} \le \frac{K\rho}{\sigma^3} \text{ a.s.},$$
$$\limsup_{n\to\infty} n^{\frac{1}{2}}(\ln n)^{-\frac{1}{2}} r^{*2}_{N_n}(x)x \le \frac{c\rho}{\sigma^3} \text{ a.s..}$$

Acknowledgements

The author owes thanks to Professor T. M. Tuan and Professor T. H. Thao for their guidance and continuous support. The author is grateful to the anonymous referee who gave a carefully and thoroughly compiled list of places where presentation could be improved.

References

1. R. Beran, *Bootstrap method in statistics*, Jahsesber. Deutsch. Math.-Verein **86** (1984), 14-30.

2. B. Efron, *Bootstrap methods: Another look at the Jackknife*, Ann. Statist. **7** (1979), 1-26.
3. B. Efron, *Nonparametric standard errors and confidence intervals (with discussion)*, Canad. J. Statist. **9** (1981), 139-172.
4. B. Efron and R. Tibshirani, *Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion)*, Statist. Sci. **1** (3) (1986), 54-47.
5. P. Hall, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York, 1992.
6. V. M. Kruglov, V. Yu. Korolev, *Limit Theorems for Random Sums* (in Russian), Moscow University Edition, 1990.
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*, Bulletin, College of Science, University of Hue, **10** (1) (1996), 48-52.
9. Nguyen Van Toan, *On the bootstrap estimate with random sample size*, Scientific Bulletin of Universities (1998), 31-34.
10. Nguyen Van Toan, *On the asymptotic accuracy of the bootstrap with random sample size*, Vietnam Journal of Mathematics **26** (4) (1998), 351-356.
11. Nguyen Van Toan, *On the asymptotic accuracy of the bootstrap with random sample size*, Pakistan Journal of Statist. **14** (3) (1998), to appear.
12. K. Singh, *On the asymptotic accuracy of Efron's bootstrap*, Ann. Statist. **9** (6) (1981), 1187-1195.
13. Tran Manh Tuan and Nguyen Van Toan, *On the asymptotic theory for the bootstrap with random sample size*, Proceed. National Centre for Science and Technology of Vietnam **10** (2) (1998), 3-8.
14. Tran Manh Tuan and Nguyen Van Toan, *An asymptotic normality theorem of the bootstrap sample with random sample size*, VNU Journal of Science, Nat. Sci. tXIV, No. 1 (1998), 1-7.

Department of Mathematics, College of Science,
Hue University