The Annals of Statistics
1987, Vol. 15, No. 2, 724–731

BOOTSTRAP OF THE MEAN IN THE INFINITE VARIANCE CASE

BY K. B. ATHREYA

Iowa State University

Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables with $EX_1^2 = \infty$ but X_1 belonging to the domain of attraction of a stable law. It is known that the sample mean $\overline{X}_n$ appropriately normalized converges to a stable law. It is shown here that the bootstrap version of the normalized mean has a random distribution (given the sample) whose limit is also a random distribution implying that the naive bootstrap could fail in the heavy tailed case.

1. Introduction. It is now nearly eight years since Efron [6] introduced the term *bootstrap* into the statistical literature to denote a variety of resampling methods. Theoretical justifications have been attempted with some success for Efron's particular bootstrap method, namely, simple random sampling (s.r.s.) with replacement from the original sample. In particular, the papers of Singh [8] and Bickel and Freedman [4] showed that if $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables (i.i.d.r.v.) with $EX_1^2 < \infty$ and if given $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$, we choose $Y_1, Y_2, \ldots, Y_n$ by a s.r.s. with replacement from the set $\{X_1, X_2, \ldots, X_n\}$ and if

$$H_n(x, \omega) \equiv P\left(\frac{\overline{Y}_n - \overline{X}_n}{s_n} \le x | \mathbf{X}_n\right),$$

where

$$\overline{Y}_n = n^{-1}\Sigma_1^n Y_i, \qquad \overline{X}_n = n^{-1}\Sigma_1^n X_i, \qquad s_n^2 = n^{-1}\Sigma_1^n \left(X_i - \overline{X}_n\right)^2$$

then $\sup_x |H_n(x, \omega) - \Phi(x)| \to 0$ w.p.1. Singh [8] showed further that $H_n(\cdot, \omega)$ is a better approximation to the distribution of $\sigma^{-1}(\overline{X}_n - \mu)$ than the Edgeworth approximation up to the first-order term. Bickel and Freedman [4] investigated the performance of the bootstrap versions of other statistics and showed that in general under second moments the bootstrap asymptotics is the same as that supplied by the normal theory.

The present paper addresses the simple question of what happens to the distribution of the bootstrap version of the sample mean when $EX_1^2 = \infty$. We investigate the case when X_1 belongs to the domain of attraction of a stable law of order α, $0 < \alpha \le 2$. We find that the bootstrap is not successful here. The limiting distributions of the sample mean and its bootstrap version are quite different, the latter one being a random probability distribution.

In Section 2, we give a precise description of this result. The proof is given in Section 3. Some remarks are made in Section 4.

Received September 1984; revised August 1986.
AMS 1980 *subject classifications*. Primary 62E, 62F; secondary 60F.
Key words and phrases. Bootstrap, stable law, Poisson random measure.

2. The result. Let $X_1, X_2, \ldots, X_n$ be i.i.d.r.v. with a distribution function $F(\cdot)$ satisfying

$$1 - F(x) \sim x^{-\alpha}L(x), \qquad F(-x) \sim cx^{-\alpha}L(x), \tag{2.1}$$

as $x \to \infty$, where $L(\cdot)$ is slowly varying at ∞ and c is a nonnegative constant. For simplicity of exposition, we *assume throughout that* $1 < \alpha < 2$. The extension of the result here to the cases $0 < \alpha \le 1$ and $\alpha = 2$ has been worked out and may be found in Athreya [1], [2].

Given $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$, let $Y_1, Y_2, \ldots, Y_n$ be i.i.d.r.v. with distribution

$$P(Y_1 = X_j | \mathbf{X}_n) = n^{-1}, \quad \text{for } j = 1, 2, \ldots, n. \tag{2.2}$$

It is known that $\overline{X}_n \equiv n^{-1}\Sigma_1^n X_i$, appropriately normalized, converges to a stable law (see Feller [7]). More precisely, let $\{a_n\}$ be an increasing sequence going to ∞ such that

$$nP(X_1 \ge a_n) = na_n^{-1}L(a_n) \to 1 \tag{2.3}$$

and let

$$\mu \equiv EX_1 \quad \text{and} \quad R_n \equiv na_n^{-1}(\overline{X}_n - \mu). \tag{2.4}$$

Then

$$\sup_x |P(R_n \le x) - G_\alpha(x)| \to 0, \tag{2.5}$$

where $G_\alpha(x)$ is the distribution function of a stable law of order α whose characteristic function is

$$\phi(t) = \exp\left(\int (e^{itx} - 1 - itx)\lambda_\alpha(dx)\right), \tag{2.6}$$

where $\lambda_\alpha(\cdot)$ is a measure on the real line such that for $x > 0$

$$\lambda_\alpha[x, \infty) = x^{-\alpha} \quad \text{and} \quad \lambda_\alpha(-\infty, -x] = cx^{-\alpha}. \tag{2.7}$$

The bootstrap version of R_n of (2.4) is

$$T_n = nX_{nn}^{-1}(\overline{Y}_n - \overline{X}_n), \tag{2.8}$$

where $X_{nn} = \max(X_1, X_2, \ldots, X_n)$. Let

$$H_n(x, \omega) \equiv P(T_n \le x | \mathbf{X}_{(n)}). \tag{2.9}$$

If the bootstrap were to be successful here, then $H_n(\cdot, \omega)$ should converge to $G_\alpha(\cdot)$ in distribution. However, this is not the case. There is a random limit. To describe this precisely, we proceed as follows. We assume without loss of generality that on the same probability space (Ω, B, P) there exists a Poisson random measure $N'(\cdot, \cdot)$, that is, a family of random variables $\{N'(A, \omega);\ A$ ranging over the Borel sets of $R\}$ such that for disjoint sets $A_1, A_2, \ldots, A_k$, $\{N'(A_i, \omega),\ i = 1, 2, \ldots, k\}$ are independent Poisson random variables with mean $\lambda_\alpha(A_i)$, $i = 1, 2, \ldots, k$, respectively, where $\lambda_\alpha(\cdot)$ is as in (2.7).

Since $EN'([x, \infty), \omega) = x^{-\alpha}$ for $x > 0$, for almost all ω there exists a $\tau(\omega) > 0$ such that $N'((\tau(\omega), \infty), \omega) = 0$. We normalize N' by τ. That is, we define

$$N(A, \omega) \equiv N'(A\tau^{-1}, \omega), \quad \text{for all Borel sets } A \tag{2.10}$$

and let

$$\phi(t, \omega) = \exp\left(\int f_t(x) N(dx, \omega)\right), \tag{2.11a}$$

where

$$f_t(x) = (e^{itx} - 1 - itx). \tag{2.11b}$$

It turns out that for each ω, $\phi(t, \omega)$ is the characteristic function of an infinitely divisible nondegenerate probability distribution function $H(x, \omega)$. (For proof of this, see Athreya [1].) For future reference, we write

$$\phi(t, \omega) = \int e^{itx} H(dx, \omega). \tag{2.12}$$

Our main result is

THEOREM 1. *For any set of real numbers* $x_1, x_2, \ldots, x_k$, *the sequence of random vectors* $(H_n(x_i, \omega), i = 1, 2, \ldots, k)$ *converges in distribution to the random vector* $(H(x_i, \omega), i = 1, 2, \ldots, k)$.

COROLLARY 1. *For any* $x_1 < x_2$, $H_n(x_2, \omega) - H_n(x_1, \omega)$ *converges in distribution to* $H(x_1, \omega) - H(x_2, \omega)$.

One could think of $H_n(\cdot, \omega)$ and $H(\cdot, \omega)$ as stochastic processes whose trajectories are in the Skorohod space $D(-\infty, +\infty)$. Theorem 1 says that H_n converges to H in the sense of finite-dimensional distributions. Professor J. K. Ghosh asked whether this could be improved to weak convergence in the Skorohod space. The answer is yes, and the proof may be found in Athreya [2].

3. Proof of Theorem 1. Let $H_n(\cdot, \cdot)$ be as in (2.9) and

$$\phi_n(t, \omega) = E\left(e^{itT_n} | \mathbf{X}_n\right) = \int e^{itx} H_n(dx, \omega). \tag{3.1}$$

Since $Y_1, Y_2, \ldots, Y_n$ are independent given $\mathbf{X}_n$,

$$\phi_n(t, \omega) = \left(1 - \psi_n(t, \omega)\right)^n, \tag{3.2}$$

$$\psi_n(t, \omega) = 1 - E\left(e^{it(Y_1 - \overline{X}_n)X_{nn}^{-1}} | \mathbf{X}_n\right). \tag{3.3}$$

Since $\Sigma_1^n(X_j - \overline{X}_n) = 0$,

$$\begin{aligned} -n\psi_n(t, \omega) &= \sum_{j=1}^{n} \left(e^{it(X_j - \overline{X}_n)X_{nn}^{-1}} - 1 - it\left(X_j - \overline{X}_n\right)X_{nn}^{-1}\right) \\ &= \int f_t(x) N_n(dx, \omega), \end{aligned} \tag{3.4}$$

where $f_t(x)$ is as in (2.11b) and $N_n(\cdot, \omega)$ is a random measure defined by

$$N_n(A, \omega) = \sum_{j=1}^{n} \chi_A\big((X_j - \overline{X}_n)X_{nn}^{-1}\big). \tag{3.5}$$

According to Theorem A of the Appendix, to prove Theorem 1 it suffices to show that for arbitrary $t_1, t_2, \dots, t_k$, the sequence of random vectors $(\phi_n(t_i, \omega), i = 1, 2, \dots, k)$ converges in distribution to $(\phi(t_i, \omega), i = 1, 2, \dots, k)$. This in turn is equivalent to showing that $(n\psi(t_i, \omega), i = 1, 2, \dots, k)$ converges in distribution to that of $(\psi(t_i, \omega), i = 1, 2, \dots, k)$, where $\psi(t, \omega) = \int f_t(x)N(dx, \omega)$ with $f_t(\cdot)$ as in (2.11b) and $N(\cdot, \cdot)$ as in (2.10). By the Cramér–Wold device, this is equivalent to showing that for arbitrary $l_1, l_2, \dots, l_k$,

$$n\sum_{1}^{K} l_j\psi_n(t_j, \omega) \equiv \int g(x)N_n(dx, \omega)$$

converges in distribution to $\int g(x)N(dx, \omega)$, where $g(x) = \sum_1^k l_j f_{t_j}(x)$ and $N(A, \omega)$ is as in (2.10). Since for fixed t, $f_t(x) = O(x^2)$ as $x \to 0$ and $= O(|x|)$ as $x \to \infty$, the real and imaginary parts of the function $g(x)$ satisfy the hypothesis of Proposition 1 below and hence the proof of Theorem 1 is completed. It remains to state and prove

PROPOSITION 1. *Let $g(\cdot)$ be a continuous real-valued function on R such that $g(x) = O(x^2)$ as $x \to 0$ and $= O(|x|)$ as $|x| \to \infty$. Let $N_n(\cdot,\cdot)$ be as in* (3.5) *and $N(\cdot,\cdot)$ be as in* (2.10). *Then*

$$\int g(x)N_n(dx, \omega) \to_d \int g(x)N(dx, \omega), \tag{3.6}$$

where $\to_d$ stands for convergence in distribution.

PROOF. By Theorem B of the Appendix, the sequence of random measures $N_n(\cdot, \omega)$ introduced in (3.5) satisfies the property that for disjoint intervals $I_1, I_2, \dots, I_k$ whose closure is contained in $R - \{0\}$,

$$\big(N_n(I_j, \omega),\ j = 1, 2, \dots, k\big) \to_d \big(N(I_j, \omega),\ j = 1, 2, \dots, k\big), \tag{3.7}$$

where $N(\cdot,\cdot)$ is as in (2.10). We now show that for each $\varepsilon > 0$

$$\lim_{\eta \downarrow 0} \lim_{n} P\big(D_{n1}(\eta, \omega) > \varepsilon\big) = 0 \tag{3.8}$$

and

$$\lim_{\zeta \uparrow \infty} \lim_{n} P\big(D_{n2}(\zeta, \omega) > \varepsilon\big) = 0, \tag{3.9}$$

where

$$D_{n1}(\eta, \omega) = \int_{|x| < \eta} x^2 N_n(dx, \omega) \tag{3.10}$$

and

$$D_{n2}(\zeta, \omega) = \int_{|x| > \zeta} |x| N_n(dx, \omega). \tag{3.11}$$

From the definition of $N_n(\cdot,\cdot)$ in (3.5) we have

$$D_{n1}(\eta,\omega) \le 2\left(\sum_j X_j^2 X_{nn}^{-2} + n\overline{X}_n^2 X_{nn}^{-2}\right),$$

where Σ extends over $j \ni |X_j| \le \eta X_{nn} + \overline{X}_n$.

By Theorem B of the Appendix, $X_{nn}^{-1}a_n$ converges in distribution. Also by the strong law $\overline{X}_n$ is bounded and by choice $na_n^{-2} \to 0$. To establish (3.8) it suffices to show that

$$\lim_{\eta \downarrow 0} \lim_n P\left(\sum_j{}' X_j^2 a_n^{-2} > \varepsilon\right) = 0, \tag{3.12}$$

where Σ' extends over $j \ni |X_j| < \eta a_n$. Now $E(\Sigma_j' X_j^2 a_n^{-2}) = na_n^{-2}E(X_1^2\colon |X_1| < \eta a_n)$, which by our hypothesis (2.1), is asymptotic to const. $na_n^{-2}(\eta a_n)^{2-\alpha}L(\eta a_n)$ (Feller [7], page 544) and hence to const. $\eta^{2-\alpha}$. Since $\alpha < 2$, (3.12) follows. Turning now to (3.9) and arguing as above, it suffices to show that

$$\lim_{\zeta \uparrow \infty} \lim_n P\left(\sum_j{}'' a_n^{-1}|X_j| > \varepsilon\right) = 0, \tag{3.13}$$

where Σ'' extends over $j \ni |X_j| > \zeta a_n$.

Now $E(\Sigma'' a_n^{-1}|X_j|) = na_n^{-1}E(|X_1|\colon |X_1| > \zeta a_n)$, which by our hypothesis (2.1), is asymptotic to const. $na_n^{-1}(\zeta a_n)^{1-\alpha}L(\zeta a_n)$ (Feller [7], page 544) and hence to const. $\zeta^{1-\alpha}$. Since $\alpha > 1$, (3.13) follows. By standard analysis (3.7) implies that for continuous functions $g(\cdot)$ with compact support in $R - \{0\}$,

$$\int g(x) N_n(dx,\omega) \to_d \int g(x) N(dx,\omega). \tag{3.14}$$

Since (3.8) and (3.9) hold, (3.14) extends to all g satisfying the conditions of Proposition 1. □

4. Remarks. What, if any, is the significance of Theorem 1? It says that if one does a naive bootstrap on the sample mean and if the underlying population does not have a finite variance then the bootstrap distribution will not converge to the same limit as the sample mean. Thus, constructing confidence intervals on the basis of a Monte Carlo simulation of the bootstrap could yield misleading results. So unless one is reasonably sure that the underlying distribution is not heavy tailed, one should hesitate to use the naive bootstrap. In particular, in variance estimation using naive bootstrap could be bad if the underlying population has no fourth moment. There are some modifications of the bootstrap method such as changing the resample size from n to m with $m = O(n)$ or trimming the sample and doing bootstrap on the reduced sample. These do lead to inefficiencies, the precise nature of which needs to be studied.

A referee has pointed out that the phenomenon of obtaining a limit that is a random measure is familiar from other examples such as the Hodges estimate (see Beran [3]) and also that the idea of changing the sample size to m is also not new. It appears in Bretagnolle [5].

APPENDIX

The following generalization of the Lévy–Cramér continuity theorem was used in the proof of Theorem 1.

THEOREM A. *Let* $\{H_n(x, \omega)\}$, $n = 0, 1, 2, \ldots,$ *be a sequence of random distribution functions on a probability space* (Ω, B, P). *That is, for each* ω *and* n, $H_n(x, \omega)$ *is a distribution function. Assume that for any* $x_1, x_2, \ldots, x_k$ *and* n, $(H_n(x_i, \omega),\ i = 1, 2, \ldots, k)$ *is a measurable map from* (Ω, B) *to* $([0,1])^k$. *Let* $\phi_n(t, \omega) = \int e^{itx} H_n(dx, \omega)$ *be the characteristic function of* H_n. *Then, for any* $(x_1, x_2, \ldots, x_k)$ *as* $n \to \infty$,

$$(\text{A.1}) \qquad \big(H_n(x_i, \omega),\ i = 1, 2, \ldots, k\big) \to_d \big(H_0(x_i, \omega),\ i = 1, 2, \ldots, k\big),$$

iff for any $(t_1, t_2, \ldots, t_k)$

$$(\text{A.2}) \qquad \big(\phi_n(t_i, \omega),\ i = 1, 2, \ldots, k\big) \to_d \big(\phi_0(t_i, \omega),\ i = 1, 2, \ldots, k\big).$$

PROOF. It can be shown by standard methods that both (A.1) and (A.2) are equivalent to

$$(\text{A.3}) \qquad \int f(x) H_n(dx, \omega) \to_d \int f(x) H_0(dx, \omega),$$

for any bounded continuous function f on R. The details of that argument may be found in Athreya [1]. □

The following result about the convergence of N_n to N was used in the proof of Theorem 1.

THEOREM B. *Let* $N_n(A, \omega)$ *be as in* (3.5) *and* $N(A, \omega)$ *be as in* (2.10). *Then for disjoint intervals* $I_1, I_2, \ldots, I_k$ *contained in* $R - \{0\}$

$$(\text{A.4}) \qquad \big(N_n(I_j, \omega),\ j = 1, 2, \ldots, k\big) \to_d \big(N(I_j, \omega),\ j = 1, 2, \ldots, k\big).$$

PROOF. Since $\{\overline{X}_n\}$ is bounded w.p.1 and $X_{nn} \to \infty$, (A.4) is implied by the result that for each k

$$(\text{A.5}) \qquad \begin{aligned} &a_n^{-1}\big(X_{nn}, X_{n(n-1)}, \ldots, X_{n(n-k+1)}, X_{nk}, X_{n(k-1)}, \ldots, X_{n1}\big) \\ &\quad \to_d \big(\tau_1, \tau_2, \ldots, \tau_k, \tau_k', \tau_{k-1}', \ldots, \tau_1'\big), \end{aligned}$$

where for each n, $X_{n1} \le X_{n2} \le \cdots \le X_{nn}$ are the order statistics of $X_1, X_2, \ldots, X_n$, and where a_n is chosen as in (2.3), and for $i = 1, 2, \ldots, k$

$$\tau_i = \sup\{x \colon x > 0,\ N'([x, \infty), \omega) = i\}$$

and

$$\tau_i' = \inf\{y \colon y < 0,\ N'((-\infty, j], \omega) = i\},$$

with $N'(\cdot, \cdot)$ being the Poisson random measure introduced in Section 2. To establish (A.5) let $s_1 > r_1 > s_2 > r_2 > \cdots > s_k > r_k > 0 > s_k' > r_k' > s_{k-1}' > r_{k-1}' > \cdots > s_1' > r_1'$ be given and let $J_j = (r_j, s_j)$ and $J_j' = (r_j', s_j')$.

Now note that the event

$$\text{(A.6)}\quad \left\{a_n^{-1}X_{n(n-j)}\varepsilon J_{j+1},\ j=0,1,2,\dots,k-1;\ a_n^{-1}X_{nj}\varepsilon J_j',\ j=1,2,\dots,k\right\}$$

is the same as the event

$$\text{(A.7)}\quad \begin{aligned}&\left\{N_n'(J_j)=1 \text{ for } j=1,2,\dots,k,\ N_n\left([r_k,\infty)-\bigcup_1^k J_j\right)=0,\right.\\ &\left.N_n'(J_j')=1 \text{ for } j=1,2,\dots,k,\ N_n\left((-\infty,s_k']-\bigcup_1^k J_j'\right)=0\right\},\end{aligned}$$

where $N_n'(A)=\sum_1^n\chi_A(a_n^{-1}X_j)$.

Since $X_1, X_2,\dots,X_n$ are i.i.d. the random vector $\{N_n'(A_j)\colon\ j=1,2,\dots,r,\ N_n'(R-\bigcup_1^r A_j)\}$ for disjoint $A_1, A_2,\dots,A_r$ has an $(r+1)$-cell multinomial distribution with parameters $(n;\ p_{nj},\ j=1,2,\dots,r,\ 1-\sum_1^r p_{nj})$, where $p_{nj}=P(a_n^{-1}X_1\varepsilon A_j)$.

By our hypothesis (2.1)

$$nP\left(a_n^{-1}X_1\varepsilon J_j\right)\to\left(r_j^{-\alpha}-s_j^{-\alpha}\right)\quad\text{and}\quad nP\left(a_n^{-1}X_1\varepsilon J_j'\right)\to c\left(|s_j'|^{-\alpha}-|r_j'|^{-\alpha}\right).$$

By the well-known convergence of the multinomial to the Poisson

$$\begin{aligned}&\lim_n P(\text{event in (A.6)})\\ &\quad=\lim_n P(\text{event in (A.7)})\\ &\quad=P\left\{N'(J_j)=1\quad \text{for } j=1,2,\dots,k,\ N'\left([r_k',\infty)-\bigcup_1^k J_j\right)=0;\right.\\ &\qquad\left.N'(J_j')=1\quad\text{for } j=1,2,\dots,k,\ N'\left((-\infty,s_k']-\bigcup_1^k I_j'\right)=0\right\}.\end{aligned}$$

It is easy to see that this is the same as (A.5). □

Acknowledgments. The author wishes to thank the Editors of the *Annals*, Professors M. Perlman and W. van Zwet, the latter for a very constructive review of an earlier version, and Professors J. K. Ghosh and D. Freedman for suggesting the problem.

REFERENCES

[1] ATHREYA, K. B. (1984). Bootstrap for the mean in the infinite variance case. Technical Report 86-22, Dept. of Statistics, Iowa State Univ.

[2] ATHREYA, K. B. (1985). Bootstrap for the mean in the infinite variance case. II. Technical Report 86-21, Dept. of Statistics, Iowa State Univ.

[3] BERAN, R. (1982). Estimated sampling distributions: the bootstrap and competitors. *Ann. Statist.* **10** 212–225.

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

[5] BRETAGNOLLE, J. (1983). Lois limites du bootstrap de certaines fonctionnelles. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **19** 281–296.

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

[7] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

[8] SINGH, K. (1981). On the asymptotic efficiency of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195.

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011