The Annals of Statistics
1989, Vol. 17, No. 3, 1168–1175

ON THE BOOTSTRAP OF THE SAMPLE MEAN IN THE INFINITE VARIANCE CASE[1]

By Keith Knight

University of Toronto

Athreya showed that the bootstrap distribution of a sum of infinite variance random variables did not (with probability 1) tend weakly to a fixed distribution but instead tended in distribution to a random distribution. In this paper, we give a different proof of Athreya's result motivated by a heuristic large sample representation of the bootstrap distribution.

1. Introduction. Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed (i.i.d.) random variables (r.v.'s) from some distribution F. Given $X_1, \ldots, X_n$, we define the empirical distribution function F_n as

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} I(X_k \le x),$$

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

We now take $X_1^*, \ldots, X_n^*$ independent with distribution F_n and consider the distribution of the sample mean, $\bar{X}^*$, of the X_k^*'s ("conditional" on $X_1, \ldots, X_n$). This is called the bootstrap distribution of the sample mean, $\bar{X}$, and, like F_n, is a random distribution in the sense that it is a function of the observations $X_1, \ldots, X_n$.

The large sample behaviour of the bootstrap distribution is well-known when $E(X_1^2) < \infty$; along almost all sample sequences, the distribution of

$$\sqrt{n}\,(\bar{X}^* - \bar{X}) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k^* - \bar{X})$$

converges weakly to a normal distribution with mean 0 and variance $\sigma^2 = \mathrm{Var}(X_1)$ [Bickel and Freedman (1981)]. More precisely, if μ_n^* denotes the bootstrap distribution of $\sqrt{n}\,(\bar{X}^* - \bar{X})$ and μ is the (fixed) normal distribution, then

$$\int f(x)\mu_n^*(dx) \to_{a.s.} \int f(x)\mu(dx) = \int f(x)(2\pi)^{-1/2}\sigma^{-1}\exp(-x^2/2\sigma^2)\,dx$$

for all bounded, continuous functions f; in other words, μ_n^* converges weakly to μ with probability 1.

Now consider a sequence r.v.'s $\{X_k\}$, which are in the domain of attraction of a stable law with infinite second moment. This means that there exist constants

Received April 1988; revised October 1988.

[1]Research supported by a grant from the Natural Sciences and Engineering Research Council of Canada and by ONR Contract N00014-84-C-0169.

AMS 1980 *subject classifications*. Primary 62E20; secondary 60B05, 60G57.

Key words and phrases. Bootstrap, stable law, random probability measures, weak convergence.

$\{a_n\}$ and $\{b_n\}$ such that

$$S_n = a_n^{-1} \sum_{k=1}^{n} (X_k - b_n) \to_d S_\alpha,$$

where S_α is a stable r.v. with index $\alpha \in (0,2)$; see Feller (1971) for more details. Athreya (1987) showed that the bootstrap distribution of the sample mean does not converge weakly to a fixed distribution (as it does in the finite variance case) but instead converges in distribution (with respect to the weak topology on the space of bounded measures) to a random probability distribution. He showed that the characteristic function of the appropriately centered and scaled bootstrap mean converges in distribution on the space of continuous complex valued functions to a limiting random characteristic function; it can be shown that convergence in distribution of characteristic functions is equivalent to convergence in distribution of the corresponding random probability measures (with respect to the weak topology). Athreya also finds a representation for the (random) characteristic function of the limiting random distribution in terms of a nonhomogeneous Poisson process. (In this paper, $\to_d$ will be used to denote convergence in distribution of random elements on several different spaces; the relevant space should be clear from the context.)

The purpose of this paper is to present a different proof of Athreya's result. This "probabilistic" proof is motivated by a heuristic large sample representation of the distribution of the bootstrap mean.

2. Results. Assume that $X_1, \dots, X_n$ are i.i.d. r.v.'s in the domain of attraction of a stable law with index $\alpha \in (0,2)$. Taking $X_1^*, \dots, X_n^*$ from the empirical distribution F_n, we consider the distribution of the normed sum

$$S_n^* = a_n^{-1} \sum_{k=1}^{n} (X_k^* - b^*), \tag{1}$$

where $b^* = \overline{X} = E^*(X_1^*)$ and the constants a_n are such that

$$nP(|X_1| > a_n x) \to x^{-\alpha} \quad \text{as } n \to \infty.$$

(All r.v.'s which are generated by bootstrap sampling are superscripted with an asterisk; P^* and E^* denote probability and expectation for such r.v.'s.) Note that if $X_1, \dots, X_n$ are regarded as fixed constants, then S_n^* has the same distribution as

$$a_n^{-1} \sum_{k=1}^{n} X_k (M_{nk}^* - 1),$$

where $(M_{n1}^*, M_{n2}^*, \dots, M_{nn}^*)$ is a multinomial random vector with n trials and each cell probability n^{-1}. It is easy to show that for large n, this random vector has approximately the same distribution as a vector of n independent Poisson r.v.'s with mean 1; more precisely

$$(M_{n1}^*, \dots, M_{nn}^*, 0, 0, \dots) \to_d (M_1^*, M_2^*, \dots),$$

where the limit is a sequence of i.i.d. Poisson r.v.'s. This suggests that for large n,

$$S_n^* \approx_d a_n^{-1} \sum_{k=1}^{n} X_k(M_k^* - 1),$$

which in turn suggests that the large sample behaviour of the distribution of S_n^* may depend on the large sample behaviour of the $a_n^{-1}X_k$'s.

We will now define notation similar to that found in Lepage, Woodroofe and Zinn (1981).

1. Let $Y_k = |X_k|$ and let $Y_{n1} \geq Y_{n2} \geq \cdots \geq Y_{nn}$ be the ordered Y_k's.
2. Let $X_{n1}, \dots, X_{nn}$ be the corresponding X values and define δ_{nk} so that $X_{nk} = \delta_{nk} Y_{nk}$.
3. Let $Z_{nk} = a_n^{-1} Y_{nk}$.
4. Let $p = \lim_{x \to \infty} P(X_1 > x)/P(|X_1| > x)$.

Using this notation, we can now state the following result, which will be important in the proof of Theorem 2.

THEOREM 1 [Lepage, Woodroofe and Zinn (1981)]. *Defining* $\{Z_{nk}\}$ *and* $\{\delta_{nk}\}$ *as above,*

(a) $\mathbf{Z}^{(n)} = (Z_{n1}, \dots, Z_{nn}, 0, 0, \dots) \to_d (Z_1, Z_2, \dots) = \mathbf{Z}$,
(b) $\delta^{(n)} = (\delta_{n1}, \dots, \delta_{nn}, 0, 0, \dots) \to_d (\delta_1, \delta_2, \dots) = \delta$,

where $\delta_1, \delta_2, \dots$ *are i.i.d. r.v.'s taking values* 1 *and* -1 *with probabilities* p *and* $1-p$, *respectively, and* $Z_k = (E_1 + \cdots + E_k)^{-1/\alpha}$ *for an i.i.d. sequence of exponential r.v.'s* $E_1, E_2, \dots$ *with mean* 1. *The limiting random sequences* $\mathbf{Z}$ *and* δ *are independent.*

Note that

$$S_n^* =_{d^*} \sum_{k=1}^{n} \delta_{nk} Z_{nk}(M_{nk}^* - 1)$$

and so Theorem 1 suggests that the bootstrap distribution of S_n^* will tend, in some sense, to the distribution of the r.v.

$$\sum_{k=1}^{\infty} \delta_k Z_k(M_k^* - 1).$$

This r.v. is well-defined with probability 1 since $E^*[\delta_k Z_k(M_k^* - 1)] = 0$ for all k and

$$\sum_{k=1}^{\infty} E^*\left[Z_k^2(M_k^* - 1)^2\right] = \sum_{k=1}^{\infty} Z_k^2 < \infty$$

with probability 1.

THEOREM 2. *Let* μ_n^* *be the random probability measure of* S_n^* *defined in* (1). *Then*

$$\mu_n^* \to_d \mu^*,$$

where μ^* *is the random probability measure of*

$$S^* = \sum_{k=1}^{\infty} \delta_k Z_k (M_k^* - 1) \tag{2}$$

and $\delta_1, \delta_2, \ldots$ *and* $Z_1, Z_2, \ldots$ *are as defined in Theorem* 1 *and* $M_1^*, M_2^*, \ldots$ *is an i.i.d. sequence of Poisson r.v.'s with mean* 1.

PROOF. We need to show that for all bounded continuous functions f,

$$E^*[f(S_n^*)] = \int f(x)\mu_n^*(dx) \to_d \int f(x)\mu^*(dx) = E^*[f(S^*)]. \tag{3}$$

Since R^∞ is a complete and separable metric space under the metric

$$d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{\infty} \frac{|x_i - y_i|}{1 + |x_i - y_i|} 2^{-i}$$

we can choose a probability space such that

$$d(\mathbf{Z}^{(n)}, \mathbf{Z}) \to 0 \quad \text{a.s.} \quad \text{and} \quad d(\delta^{(n)}, \delta) \to 0 \quad \text{a.s.}$$

We will show that on this probability space

$$\int f(x)\mu_n^*(dx) \to \int f(x)\mu^*(dx) \quad \text{a.s.}$$

in which case (3) follows in general. Since $Z_k \to 0$ as $k \to \infty$, it follows that

$$\sum_{k=1}^{\infty} \delta_{nk} Z_{nk} I[Z_{nk} > \varepsilon](M_{nk}^* - 1) \to_{d^*} \sum_{k=1}^{\infty} \delta_k Z_k I[Z_k > \varepsilon](M_k^* - 1),$$

where $\to_{d^*}$ indicates that the convergence in distribution occurs with respect to P^*-probability (with P-probability 1). In addition,

$$\begin{aligned}
E^*&\left[\left(\sum_{k=1}^{\infty} \delta_{nk} Z_{nk} I[Z_{nk} \le \varepsilon](M_{nk}^* - 1)\right)^2\right] \\
&= \sum_{k=1}^{n} Z_{nk}^2 I[Z_{nk} \le \varepsilon] \operatorname{Var}^*(M_{nk}^*) \\
&\quad + 2\sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \delta_{nj}\delta_{nk} Z_{nj} Z_{nk} I[Z_{nk} \le \varepsilon] \operatorname{Cov}^*(M_{nj}^*, M_{nk}^*) \\
&\le \sum_{k=1}^{n} Z_{nk}^2 I[Z_{nk} \le \varepsilon] + \frac{2}{n} \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} Z_{nj} Z_{nk} I[Z_{nk} \le \varepsilon] \\
&\le 2 \sum_{k=1}^{n} Z_{nk}^2 I[Z_{nk} \le \varepsilon] \\
&\to 2 \sum_{k=1}^{\infty} Z_k^2 I[Z_k \le \varepsilon] \quad \text{a.s. as } n \to \infty \\
&\to 0 \quad \text{a.s. as } \varepsilon \to 0.
\end{aligned}$$

Hence

$$\lim_{\varepsilon \to 0} \limsup_{n \to \infty} P^* \left[\left| \sum_{k=1}^{\infty} \delta_{nk} Z_{nk} I[Z_{nk} \le \varepsilon](M_{nk}^* - 1) \right| > \delta \right] = 0$$

with P-probability 1. Finally, it suffices to show that, with P-probability 1,

$$\lim_{\varepsilon \to 0} P^* \left[\left| \sum_{k=1}^{\infty} \delta_k Z_k I[Z_k \le \varepsilon](M_k^* - 1) \right| > \delta \right] = 0.$$

This follows since

$$E^* \left[\left(\sum_{k=1}^{\infty} \delta_k Z_k I[Z_k \le \varepsilon](M_k^* - 1) \right)^2 \right] = \sum_{k=1}^{\infty} Z_k^2 I[Z_k \le \varepsilon]$$

$$\to 0 \quad \text{a.s. as } \varepsilon \to 0. \qquad \square$$

3. Comments. (a) The characteristic function of the r.v. S^* is simply

$$E^*[\exp(itS^*)] = \exp\left[\sum_{k=1}^{\infty} (\exp(it\delta_k Z_k) - it\delta_k Z_k - 1) \right]$$

$$= \exp\left[\int_{-\infty}^{\infty} (\exp(itx) - itx - 1)\, d\Lambda(x) \right],$$

where $\Lambda(*)$ is a Poisson process with intensity

$$E(d\Lambda(x)) = \alpha |x|^{-\alpha-1}[(1-p)I(x<0) + pI(x>0)]\, dx.$$

(b) Athreya (1987) does not consider S_n^* but rather

$$T_n^* = Y_{n1}^{-1} \sum_{k=1}^{n} (X_k^* - b^*),$$

where Y_{n1}, the maximum of the $|X_k|$'s takes the place of a_n. It is easy to show using the same techniques as in the proof of Theorem 2 that the random probability measure of T_n^* tends in distribution to that of

$$T^* = Z_1^{-1} \sum_{k=1}^{\infty} \delta_k Z_k (M_k^* - 1). \tag{4}$$

More generally, we can replace a_n and b^* in the definition of S^* by other constants or r.v.'s and obtain similar results.

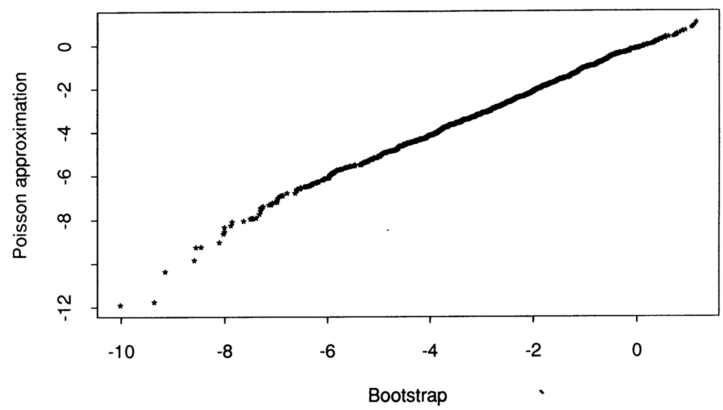

FIG. 1. *QQ plot for $n = 25$.*

(c) Theorem 2 suggests that for a given sample, $X_1, \dots, X_n$, the distribution of the bootstrap mean, $\bar{X}^*$, can be approximated by the distribution of

$$\frac{1}{n} \sum_{k=1}^{n} X_k M_k^*,$$

where $M_1^*, \dots, M_n^*$ are i.i.d. Poisson r.v.'s. Figures 1 and 2 illustrate this Poisson approximation to the bootstrap distribution for Cauchy ($\alpha = 1$) samples of size 25 and 100. For each sample, 1000 independent replications of the bootstrap mean and its Poisson approximation were made; these figures are "empirical" quantile–quantile plots of the two distributions for $n = 25$ and 100.

(d) It is possible to show that the distribution functions corresponding to S^* and T^* are continuous with probability 1. This can be done by looking at the

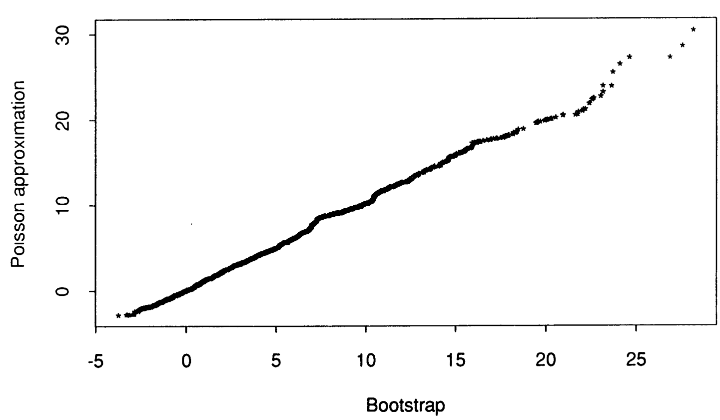

FIG. 2. *QQ plot for $n = 100$.*

characteristic functions of S^* and T^* but we will use the representations (2) and (4) and the following theorem due to Lévy.

THEOREM 3 [Lévy (1931)]. *Suppose that* $Y_1, Y_2, \dots$ *are independent r.v.'s and* $S = \sum_{k=1}^{\infty} Y_k$ *is finite. If there exists no sequence of constants* $\{\alpha_k\}$ *with*

$$\sum_{k=1}^{\infty} P(Y_k \neq \alpha_k) < \infty,$$

then S has a continuous distribution.

THEOREM 4. *Let* S^* *and* T^* *be the r.v.'s defined in* (2) *and* (4). *Both* S^* *and* T^* *have continuous distributions with probability* 1.

PROOF. Note that if Y is a Poisson r.v. with mean 1, then for any x,

$$P(Y \neq x) \geq 1 - \exp(-1).$$

By Theorem 3, it follows that the distributions can only be discontinuous if $\delta_k Z_k = 0$ all but finitely often. However, this latter event occurs with probability 0. □

(e) The phenomenon of convergence in distribution of bootstrap probability measures occurs in other situations, for example, when bootstrapping the distribution of the sample maximum. Consider the following example from Bickel and Freedman (1981), page 1210. Let $X_{(1)} > \cdots > X_{(n)}$ be the order statistics of an i.i.d. sample from a uniform distribution on the interval $(0, \theta)$ and consider the bootstrap distribution of

$$U_n^* = n(X_{(1)} - X_{(1)}^*)/X_{(1)},$$

where $X_{(1)}^*$ is the maximum of the bootstrap sample. Ideally, U_n^* should have the same limiting distribution as

$$U_n = n(\theta - X_{(1)})/X_{(1)},$$

which has a limiting exponential distribution. However,

$$P^*(X_{(1)}^* = X_{(k)}) = \left(1 - \frac{k-1}{n}\right)^n - \left(1 - \frac{k}{n}\right)^n$$
$$\to \exp(-k+1) - \exp(-k) = p_k$$

and

$$\left(n(X_{(1)} - X_{(2)})/X_{(1)}, n(X_{(1)} - X_{(3)})/X_{(1)}, \dots\right)$$
$$\to_d (E_1, E_1 + E_2, E_1 + E_2 + E_3, \dots),$$

where $E_1, E_2, \dots$ are i.i.d. exponential r.v.'s with mean 1. Thus the distribution of U_n^* tends in distribution to the distribution of U^*, where $U^* = 0$ with probability p_1 and $U^* = E_1 + \dots + E_k$ with probability p_{k+1} for $k = 1, 2, \dots$.

Acknowledgments. The author would like to thank Jon Wellner for suggesting the approach used to prove Theorem 2. The original draft of this paper was written while the author was at the University of British Columbia.

REFERENCES

ATHREYA, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Ann. Statist.* **15** 724–731.

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

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

LEPAGE, R., WOODROOFE, M. and ZINN, J. (1981). Convergence to a stable distribution via order statistics. *Ann. Probab.* **9** 624–632.

LÉVY, P. (1931). Sur les séries dont les termes sont des variables êventuelles indépendantes. *Studia Math.* **3** 119–155.

DEPARTMENT OF STATISTICS
UNIVERSITY OF TORONTO
100 ST. GEORGE STREET
TORONTO, ONTARIO
CANADA M5S 1A1