The Annals of Statistics
1983, Vol. 11, No. 3, 999–1003

INFERENCE ON MEANS USING THE BOOTSTRAP

BY G. JOGESH BABU[1] AND KESAR SINGH[2]

Rutgers University

We study the asymptotic accuracy of the bootstrap approximation to the distribution of a k-sample studentized mean.

1. Introduction and main results. Let $F_1, F_2, \cdots, F_k$ be the distributions of k populations with means $\mu_1, \mu_2, \cdots, \mu_k$. Let $\theta = \sum_1^k l_i\mu_i$ where $l_1, l_2, \cdots, l_k$ are non-zero constants. Let $\{X_{i1}, X_{i2}, \cdots, X_{in_i}\}$, $i = 1, 2, \cdots, k$, be independent random samples of sizes $n_1, n_2, \cdots, n_k$ from $F_1, F_2, \cdots, F_k$. Let n denote the vector $(n_1, n_2, \cdots, n_k)$ and $N = \sum_1^k n_i$. A natural estimator for θ is $\hat{\theta}_n = \sum_1^k l_i\bar{X}_i$ and a consistent estimator for its variance is $v_n^2 = \sum_1^k l_i^2 s_i^2/n_i$ where $\bar{X}_i = n_i^{-1}\sum_{j=1}^{n_i} X_{ij}$ and $s_i^2 = (1/n_i)\sum_1^{n_i}(X_{ij} - \bar{X}_i)^2$. Here we study the accuracy of the bootstrap approximation to the distribution of the studentized random variable $t_n = (\hat{\theta}_n - \theta)/v_n$. This approximation is discussed in the next paragraph. Although one could base an inference about θ on the difference $\hat{\theta}_n - \theta$ itself, it turns out that the bootstrap approximation is asymptotically more accurate for t_n than for $(\hat{\theta}_n - \theta)$.

Let G_i denote the empirical distribution function based on $\{X_{i1}, X_{i2}, \cdots X_{i,n_i}\}$, $i = 1, 2, \cdots, k$. The dependence of G_i's on the sample sizes is suppressed in the notation. Now let $(Y_{i1}, Y_{i2}, \cdots, Y_{in_i})$, $i = 1, 2, \cdots, k$, denote independent random samples from the populations $G_1, G_2, \cdots, G_k$; $\bar{Y}_i = n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}$ and $\gamma_i^2 = n_i^{-1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2$. Then, by definition, the distribution of $t_n^* = \sum_1^k l_i(\bar{Y}_i - \bar{X}_i)/\sum_1^k l_i^2\gamma_i^2$ under $G_1, G_2, \cdots, G_k$ is the bootstrap distribution of t_n^*. Under the conditions given below, the bootstrap distribution of t_n^* is shown to be asymptotically close to the actual distribution of t_n up to $o(N^{-1/2})$. In applications the bootstrap distribution is approximated by drawing samples of sizes $n_1, n_2, \cdots, n_k$ from $G_1, G_2, \cdots, G_k$ a large number of times, say M times, calculating t_n^* each time and finally forming an empirical histogram. It is shown here that this second stage approximation is good up to $o(N^{-1/2})$ provided $M/(N \log N) \to \infty$.

We now state the main results proved in this note. Throughout, we make the following assumptions, to be referred to as A in the sequel.

A. F_i has finite 6th moment for all $1 \le i \le k$. For at least one i, F_i is continuous. Without loss of generality we shall assume that F_1 is continuous. The n_i's tend to infinity at the same rate. In other words, the $N/n_i \le \lambda < \infty$ for all $i = 1, 2 \cdots, k$. In practice this last condition means that the n_i's are of comparable size.

In what follows, for any distribution F, let $F^{-1}(t) = \inf\{x: F(x): \ge t\}$, where $0 < t < 1$.

THEOREM. *If H_n denotes the* d.f. *of t_n and H_n^* denotes the* d.f. *of t_n^* then, under* A, *as* $N \to \infty$

$$N^{1/2}\sup_{x\in R}|H_n(x) - H_n^*(x)| \to 0 \tag{1}$$

and

$$N^{1/2}|H_n^{-1}(t) - H_n^{*-1}(t)| \to 0 \tag{2}$$

a.s. *for all $t \in (0, 1)$. Further let $H_{n,M}$ denote the approximation to H_n^* described in the second paragraph above with M samples from G_i's. If $M/(N\log N) \to \infty$ as $N \to \infty$, then*

Received November 1981; revised August 1982.

[1] On leave from the Indian Statistical Institute.

[2] Research supported by NSF Grant MCS-81-02341.

AMS 1970 *subject classifications.* 62G05; 62G15.

Key words and phrases. Bootstrap, Edgeworth expansions, empirical distribution function, confidence interval.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

for almost all sample sequences $\{X_{ij}\}$

(3) $$N^{1/2}\sup_{x\in R}|H_{n,M}(x) - H_n^*(x)| \to 0$$

and

(4) $$N^{1/2}|H_{n,M}^{-1}(t) - H_n^{*-1}(t)| \to 0$$

a.s. *for all* $t \in (0, 1)$ *as* $N \to \infty$. *The* a.s. *here refers to the random mechanism generating the samples from* G_i's. (*We assume that all the second stage sample sequences are defined on the same space.*)

It may be mentioned here that (1) in the above theorem is an extension of (1.5) in [8] which is a result involving $(\bar{X} - \mu)/\sigma$. For constructing a confidence interval for θ, one may replace an actual quantile $H_n^{-1}(\alpha)$ of t_n by its bootstrap approximation $H_{n,M}^{-1}(\alpha)$. This approximation in the one sample case has been investigated by Efron [6] on simulated data from an asymmetric population. The procedure performed quite well (see Table 5 of [6]).

2. Proof of the theorem. We first develop some notation. Let ϕ_Σ, Φ_Σ denote the density and the d.f. of a normal variable with mean zero and dispersion matrix Σ; let ϕ, Φ denote the density and the d.f. of a standard normal variable in R; let c denote a constant, the later may denote different constants at different places. For non-negative integral vectors $\beta = (\beta_1, \cdots, \beta_r)$ and $\mathbf{x} \in R^r$ let $x^\beta = \prod_{j=1}^r x_j^{\beta_j}$, $\beta! = \beta_1!\,\beta_2! \cdots \beta_r!$, $|\beta| = \beta_1 + \cdots + \beta_r$ and $D^\beta = D_1^{\beta_1} \cdots D_r^{\beta_r}$, where $D_i^{\beta_i}$ denotes the β_ith order derivative with respect to the ith variable. Finally let $\|\mathbf{x}\|^2 = x_1^2 + \cdots x_r^2$ and $\mathbf{x} \cdot \mathbf{y} = x_1y_1 + \cdots x_ry_r$ where $\mathbf{x} = (x_1, \cdots, x_r)$ and $\mathbf{y} = (y_1, \cdots y_r)$.

We shall show that

(5) $$P(t_n^* \le x) = \Phi(x) + N^{-1/2}\int_{-\infty}^x d(y)\phi(y)\,dy + o(N^{-1/2}) \quad \text{a.s.}$$

where d is a polynomial whose coefficients depend upon F_i's. The same steps will also yield

(6) $$P(t_n \le x) = \Phi(x) + N^{-1/2}\int_{-\infty}^x d(y)\phi(y)\,dy + o(N^{-1/2}).$$

Clearly (5) and (6) imply (1). Before proving (5) we shall deduce (2), (3) and (4) from (5) and (6).

To prove (3), first note that in distribution (given the original sample) $H_{n,M}$ is the same as the empirical d.f. of $H_n^{-1}(U_i)$ where $U_1, U_2, \cdots U_M$ are i.i.d. $U[0, 1]$ random variables. If E_M denotes the empirical d.f. of $U_1, \cdots U_M$ then $\#(H_n^{-1}(U_i) \le x) = ME_M(H_n(x))$. Hence, using a well known bound on E_m, we have

$$P(\sup_{x\in R}|H_{n,M}(x) - H_n(x)| \ge 4M^{-1/2}(\log M)^{1/2})$$
$$\le P(\sup_{t\in[0,1]}|E_M(t) - t| \ge 4M^{-1/2}(\log M)^{1/2}) = O(M^{-2}).$$

Consequently, in view of Borel-Cantelli lemma,

$$\limsup_{M\to\infty} M^{1/2}(\log M)^{-1/2}\sup_{x\in R}|H_{n,M}(x) - H_n(x)| \le 4 \quad \text{a.s.}$$

Here dependence of M on N is suppressed. The claim (3) in the theorem follows from this, since $M/(N\log N) \to \infty$.

The claims (2) and (4) on quantiles follow using Lemma 1 given below which is an easy consequence of Taylor's expansion.

LEMMA 1. *Let* L_N *be a sequence of* d.f.'s *on the real line such that, for a polynomial*

a_N with its coefficients bounded in N,

$$L_N(x) = \int_{-\infty}^{x} [1 + N^{-1/2} a_N(y)]\phi(y)\, dy + o(N^{-1/2})$$

uniformly in x. Then for each $\alpha \in (0, 1)$,

$$L_N^{-1}(\alpha) = z - (\phi(z)\sqrt{N})^{-1} \int_{-\infty}^{z} a_N(y)\phi(y)\, dy + o(N^{-1/2})$$

where $z = \phi^{-1}(\alpha)$.

The proof of (5) is based on Lemmas 2–5 that follow. *In the proofs below we assume* w.l.g. *that $l_1 = l_2 = \cdots l_k = 1$.*

All proofs are given for a single sequence of realizations of $\{X_{ij}\}$ for which G_j converges weakly to F_j and $\int x^6\, dG_j \to \int x^6\, dF_j$ for $j = 1, 2, \cdots k$. Thus in view of A the results hold a.s.

LEMMA 2. *Let Y be a random vector in R^2 with mean zero and dispersion matrix $V = ((v_{ij}))$. Suppose for some $b > 1$, $\max(|v_{11}|, |v_{12}|, |v_{22}|, E\|Y\|^3) < b$. Let $a > 2$ be such that $\Delta(a) < 1/10$, where $\Delta(a) = (1/a) + E(\|Y\|^3 I(\|Y\| > a))$. Then for all $\|t\| \le a^{-2}\sqrt{N}$ and all non-negative integral vectors α, with $|\alpha| \le 3$,*

$$|D^{\alpha}(g^N(t/\sqrt{N}) - (1-(i/6\sqrt{N})E(t \cdot Y)^3)\exp(-t'Vt/2))|$$
$$\le cb^9(\Delta(a) + N^{-1/2})N^{-1/2}(\|t\|^8 + 1)\exp(-t'Vt/2 + ca^{-1}b^3\|t\|^2)$$

where for $t \in R^2$, $g(t) = E(\exp(it \cdot Y))$.

The proof is similar to the proof of Theorem 9.9 of [3].

LEMMA 3. *Suppose A holds. Let $\lambda_j = (N/n_j)^{1/2}$, $Z_j = [\lambda_j(Y_{j1} - \bar{X}_j), \lambda_j^3((Y_{j1} - \bar{X}_j)^2 - s_j^2)]$, g_j denote the characteristic function of $Z_j/\sqrt{n_j}$, B_j denote the dispersion matrix of Z_j and $B = \sum_{j=1}^{k} B_j$. Then for any $\eta > 0$,*

$$\max_{|\beta|\le 3} \int_{\|t\|\le\eta\sqrt{N}} \left| D^{\beta}\left[\prod_{j=1}^{k} g_j^{n_j}(t) - e^{-t'Bt/2}\left(1 - \frac{i}{6\sqrt{N}}\sum_{j=1}^{k}\lambda_j E(t \cdot Z_j)^3\right)\right]\right| dt = o(N^{-1/2}).$$

PROOF. Define

$$f_j(t) = (1 - (i/6\sqrt{n_j})E(t \cdot Z_j)^3)\exp(-t'B_j t/2).$$

First note that

$$\max_{|\beta|\le 3}|D^{\beta}(e^{-t'Bt/2}(1 - (i/6\sqrt{N})\textstyle\sum_{j=1}^{k}\lambda_j E(t \cdot Z_j)^3) - \prod_{j=1}^{k} f_j(t))| = O(N^{-1}(\|t\|^{3(k+1)} + 1)\exp(-t'Bt/2), \tag{7}$$

and for any non-empty subset J of $\{1, 2, \cdots, k\}$,

$$\max_{|\beta|\le 3}|D^{\beta}\textstyle\prod_{j\in J} g_j^{n_j}(t)| \le \max_{|\beta|\le 3} E\|\sum_{j\in J} n_j^{-1/2}\sum_{i=1}^{n_j} Z_{ji}\|^{|\beta|}$$
$$\le 1 + k^3 \max_{1\le j\le k} n_j^{-3/2} E\|\textstyle\sum_{i=1}^{n_j} Z_{ji}\|^3 = O(1), \tag{8}$$

where Z_{ji} are independent copies of Z_j. The last inequality above follows from the proof of Lemma 14.7 of [3] as $\sup E\|Z_j\|^3 \le b < \infty$ from some $b > 0$. Also for $1 \le j \le k$

$$\max_{|\beta|\le 3}|D^{\beta}\textstyle\prod_{i\le j} f_i(t)| = O((1 + \|t\|^{6k})\exp(-t'Bt/2)). \tag{9}$$

By (8), (9) and Lemma 2, we have for any $a > 2$, $|\beta| \le 3$ and $\|t\| \le \lambda^{-1}a^{-2}\sqrt{N}$,

$$
\begin{aligned}
&|D^{\beta}(\prod_{j=1}^{k} g_j^{n_j}(t) - \prod_{j=1}^{k} f_j(t))| \\
(10) \qquad &\le \sum_{j=1}^{k} |D^{\beta}((\prod_{i<j} f_i(t) \prod_{i>j} g_i^{n_i}(t))(g_j^{n_j}(t) - f_j(t)))| \\
&= O(N^{-1/2}(r(a) + N^{-1/2})(1 + \|t\|^{8+6k}) \exp(-t'B_1 t/2 + ca^{-1}\|t\|^2)),
\end{aligned}
$$

where $r(a) = 1/a + \sup_j E(\|Z_j\|^3 I(\|Z_j\| > a))$. It now follows from (7) and (10) that, for a $|\beta| \le 3$, the integral in Lemma 3 over $\|t\| \le \lambda^{-1}a^{-2}\sqrt{N}$ is $O(r(a)N^{-1/2}) + O(N^{-1})$.

Since F_1 is continuous, the dispersion matrix of $\mathbf{X} = (X_{11}, (X_{11} - \mu_1)^2)$ is positive definite and the c.f. h of $\mathbf{X}$ satisfies the condition $|h(t)| < 1$ for all $t \ne 0$. As a result of this and the fact that weak convergence implies convergence of c.f's. uniformly over compact sets, it follows that

$$\sup\{|g_1(t)|; \|t\| \in [\lambda^{-1}a^{-2}\sqrt{N}, \eta\sqrt{N}]\} \le \delta < 1$$

for all large N. Also,

$$\inf\{(t'B_1 t)/\|t\|^2; t \ne 0\} \ge b > 0$$

for all large N under A. Finally for any $|\beta| \le 3$

$$|D^{\beta}(g_1^{n_1}(t))| \le n_1^{|\beta|} E\|Z_1 n_1^{-1/2}\|^{|\beta|} |g_1(t)|^{n_1 - |\beta|} \le c\, N^3 |g_1(t)|^{n_1 - 3}.$$

Thus for $|\beta| \le 3$, the intergral in the lemma over $\lambda^{-1}a^{-2}\sqrt{N} \le \|t\| \le \eta\sqrt{N}$ is $O(N^{-1})$. The claim now follows by letting $a \to \infty$.

Next, an inversion theorem is obtained by combining a modification of Lemma 5 in [9] with Lemma 11.6 of [3]. The proof is deleted.

LEMMA 4. *Let P be a probability on R^k and Q denote a measure with density $[1 + N^{-1/2}p(y)]\phi_{\Sigma}(y)$ where $p(y)$ is a polynomial and Σ is a positive definite matrix of order $k \times k$. Let the coefficients of $p(y)$, $\lambda_{\max}$ and $\lambda_{\min}^{-1}$ be bounded by $M > 0$ where $\lambda_{\max}$ and $\lambda_{\min}$ denote the maximum and minimum eigen values of Σ. Then for any $\varepsilon > 0$*

$$
\begin{aligned}
|P(C) - Q(C)| \le c(k) \max_{|\beta| \le k+1} \int_{\|t\| \le c\varepsilon^{-1}\sqrt{N}} |D^{\beta}(\hat{P}(t) - \hat{Q}(t))|\, dt \\
+ c(M)[\Phi_{\Sigma}((\partial C)^{\varepsilon/\sqrt{N}}) + O(N^{-1})].
\end{aligned}
$$

Here $\hat{P}$ and $\hat{Q}$ stand for c.f.'s *of P and Q; $(\partial C)^{\varepsilon/\sqrt{N}}$ is the $\varepsilon/\sqrt{N}$ neighborhood of the boundary of C.*

Finally Lemma 5 justifies converting a multivariate one-term Edgeworth expansion into an univariate one. This result is a modification of Lemma 2.1 of [2]. A proof for the present version is contained in [1].

LEMMA 5. *Let $t = (t_1, t_2, \cdots, t_r)$ be a vector, $L = \{L_{ij}\}$ be a $r \times r$ matrix and q be a polynomial in r variables. Let $M \ge \max\{|v_{ij}|, |u_{ij}|, |t_i|, |L_{ij}|, |c_\nu|\}$, where $V = ((v_{ij}))$ is a positive definite matrix $((u_{ij})) = V^{-1}$ and c_ν are the coefficients of q. Let $|t_r| > t_0 > 0$. Then there exists a polynomial p in one variable, whose coefficients are continuous functions of t_i, L_{ij}, v_{ij}, u_{ij} and c_ν such that*

$$\int_{\{z: t\cdot z + N^{-1/2}z'Lz < u\sqrt{t'Vt}\}} (1 + N^{-1/2}q(z))\phi_V(z)\, dz = \int_{-\infty}^{u} (1 + N^{-1/2}p(y))\phi(y)\, dy + o(N^{-1/2})$$

where the $o(\cdot)$ term depends on M and t_0 only.

We now briefly sketch the proof of (5) using the lemmas. Define, $\xi_j = n_j^{-1}\sum_{i=1}^{n_j} (Y_{ji} - \bar{X}_j)^2 - s_j^2$ and $s^2 = \sum_1^k s_j^2/n_j$. From Lemmas 3 and 4 it follows that for a measurable C and

$\varepsilon > 0$,

(11)
$$P[\{\sqrt{N}\sum_1^k(\bar{Y}_j-\bar{X}_j),\ N^{3/2}\sum_1^k(\xi_j/n_j)\}\in C] = \int_C \phi_B(x)[1+N^{-1/2}a_N(x)]\,dx + o(N^{-1/2}) + O(\Phi_B(\partial B)^{\varepsilon/\sqrt{N}})$$

where a_N is a polynomial whose coefficients are polynomials in $\{\lambda_j\}$, and the moments of G_j of order 6 or less. Note that $B = \{b_{ij}\}_{2\times 2}$ with $b_{11} = Ns^2$, the variance of $\sqrt{N}\sum_1^k \bar{Y}_j$. Now (11) combined with Lemma 5 entails

(12)
$$P(s^{-1}\sum_1^k(\bar{Y}_j-\bar{X}_j)[1-(\tfrac{1}{2})s^{-2}\sum_1^k(\xi_j/n_j)]\le x) = \Phi(x) + N^{-1/2}\int_{-\infty}^{x} b(y)\phi(y)\,dy + o(N^{-1/2}),$$

where b is a polynomial whose coefficients are continuous functions of B^{-1}, λ_j and the moments of G_j of order 6 or less.

Define $C_N = \{\sqrt{N}\sum_1^k|\bar{Y}_j-\bar{X}_j| < \log N\}$ and $D_N = \{N^{3/2}|\sum_1^k(\xi_j/n_j)| < \log N\}$. On $C_N \cap D_N$ one has

(13)
$$t_n^* = s^{-1}\sum_1^k(\bar{Y}_j-\bar{X}_j)[1-(\tfrac{1}{2})s^{-2}\sum_1^k(\xi_j/n_j)] + O(N^{-1}(\log N)^2)$$

(taking $l_1 = \cdots l_k = 1$). Since the 6th moments of $\{Y_{1,j}\}$ are bounded, it follows from the proof of Theorem 2 of [7] that

(14)
$$[1-P(C_n)] + [1-P(D_n)] = o(N^{-1/2}).$$

Thus (12), (13) and (14) yield (5)

REFERENCES

[1] Babu, G. J. and Singh, K. (1981). On one term Edgeworth correction by Efron's bootstrap. Unpublished manuscript.

[2] Bhattacharya, R.N. and Ghosh, J. K. (1978). On the validity of formal Edgeworth expansion. *Ann. Statist.* **6** 435–451.

[3] Bhattacharya, R. N. and Ranga Rao, R. (1976). *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

[4] Bickel, P. J. and Freedman, D. (1981). Some asymptotics on the bootstrap. *Ann. Statist.* **9** 1196–1217.

[5] Efron, B. (1979). Bootstrap—another look at Jackknife. *Ann. Statist.* **7** 1–26.

[6] Efron B. (1981). Nonparametric standard errors and confidence intervals. Stanford Technical Report No. 67, April 1981.

[7] Michel, R. (1976). Non-uniform central limit bounds with applications to probabilities of deviations. *Ann. Probab.* **4** 102–106.

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

[9] Sweeting, T. J. (1977). Speed of convergence for the multidimensional central limit theorem. *Ann. Probab.* **5** 28–41.

Math. Stat. Division
Indian Statistical Institute
203 B. T. Road
Calcutta—700035
India

Department of Statistics
Hill Center for the Mathematical Sciences
Rutgers University
New Brunswick, New Jersey 08903