The Annals of Statistics
1991, Vol. 19, No. 2, 1062–1072

BAHADUR REPRESENTATIONS FOR UNIFORM RESAMPLING AND IMPORTANCE RESAMPLING, WITH APPLICATIONS TO ASYMPTOTIC RELATIVE EFFICIENCY

By Peter Hall

Australian National University

We derive Bahadur-type representations for quantile estimates obtained from two different types of nonparametric bootstrap resampling—the commonly used uniform resampling method, where each sample value is drawn with the same probability, and importance resampling, where different sample values are assigned different resampling weights. These results are applied to obtain the relative efficiency of uniform resampling and importance resampling and to derive exact convergence rates, both weakly and strongly, for either type of resampling.

1. Introduction. In problems involving confidence intervals and hypothesis tests, attention is commonly focused on the distribution of the statistic $T = n^{1/2}\hat{\sigma}^{-1}(\hat{\theta} - \theta)$, where $\hat{\theta}$ is an estimator of an unknown parameter θ and $n^{-1}\hat{\sigma}^2$ estimates the variance of $\hat{\theta}$. Our aim in this paper is to describe Bahadur-type representations ([1], [6], page 91ff) for bootstrap estimates of the quantiles of T, obtained from two different types of nonparametric bootstrap resampling—the commonly used uniform resampling method, where each sample value is drawn with the same probability; and importance resampling (Johns [5]), where different sample values are assigned different resampling weights. Our results lead to a concise and rigorous account of the efficiency of uniform resampling relative to importance resampling.

Our main results are described in Sections 2.2 and 2.3, dealing with the cases of uniform resampling and importance resampling respectively. Section 3 presents an outline of the proofs of the main theorems.

All our results are framed for the case of the percentile-t bootstrap. However, all the results have direct analogs for the ordinary percentile bootstrap. In particular, the following properties are identical in percentile and percentile-t cases: asymptotic variances, asymptotic efficiency of uniform and importance resampling, optimal choice of the importance sampling parameter and the order of magnitude of the difference between a bootstrap quantile estimate based on B simulations and its counterpart based on an infinite number of simulations.

2. Quantile estimation

2.1. *Notation.* Let $X, X_1, X_2, \ldots$ be independent and identically distributed d-vectors with mean μ, and assume that the quantity θ which we

Received September 1988; revised June 1990.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62G30.

Key words and phrases. Bahadur representation, bootstrap, efficiency, importance sampling, quantile, strong approximation, uniform resampling.

wish to estimate may be written as $\theta(\mu)$, a smooth function of the mean. Let $\bar{X} = n^{-1}\sum_{i \le n} X_i$ denote the mean of the n-sample $\mathscr{X} = \{X_1, \dots, X_n\}$, and take $\theta(\bar{X})$ as our estimate of $\theta(\mu)$. Denote components of vectors by superscripts, and write $\theta_i(x) = (\partial/\partial x^i)\theta(x)$ for the first partial derivative of θ with respect to x^i. The asymptotic variance of $n^{1/2}\{\theta(\bar{X}) - \theta(\mu)\}$ is given by

$$\sigma^2 = \sum_{i=1}^{d} \sum_{j=1}^{d} \theta_i(\mu)\theta_j(\mu)\sigma^{ij},$$

where $\sigma^{ij} = E\{(X - \mu)^i (X - \mu)^j\}$. To avoid trivialities, we assume that $\sigma^2 > 0$. As our estimate of σ^2 we take

$$\hat{\sigma}^2 = \sum_{i=1}^{d} \sum_{j=1}^{d} \theta_i(\bar{X})\theta_j(\bar{X})\hat{\sigma}^{ij},$$

where $\hat{\sigma}^{ij} = n^{-1}\sum_{k \le n}(X_k - \bar{X})^i (X_k - \bar{X})^j$.

Note that we may write $\hat{\sigma}^{ij} = n^{-1}\sum_{k \le n} X_k^i X_k^j - \bar{X}^i \bar{X}^j$, which is a function of three univariate means. We shall assume that $\bar{X}$, and so also the vectors X_k, have been lengthened so that they include all the components needed to compute $\hat{\sigma}$. In particular, X_k must include those products $X_k^i X_k^j$ for which $\theta_i(x)\theta_j(x)$ is not identically 0. For example, in the case where $\theta(\bar{X}) = \bar{X}^1$ is a univariate sample mean, the variance estimate is $\hat{\sigma}^2 = n^{-1}\sum_k (X_k^1)^2 - (\bar{X}^1)^2$. Here X_k should be taken as the 2-vector $(X_k^1, (X_k^1)^2)^T$. Of course, this convention means that $\theta(\bar{X})$ is a nontrivial function of only some of the components of $\bar{X}$. The extra, adjoined components are present so that we may write $\hat{\sigma} = \gamma(\bar{X})$, for a smooth function γ.

Let $X_1^*, \dots, X_n^*$ denote a resample drawn randomly, with replacement, from the sample $\mathscr{X}$. Define $\bar{X}^* = n^{-1}\sum_k X_k^*$:

$$\hat{\sigma}^{ij*} = n^{-1} \sum_{k=1}^{n} \left(X_k^* - \bar{X}^*\right)^i \left(X_k^* - \bar{X}^*\right)^j,$$

$$\hat{\sigma}^{*2} = \sum_{i=1}^{d} \sum_{j=1}^{d} \theta_i(\bar{X}^*)\theta_j(\bar{X}^*)\hat{\sigma}^{ij*}.$$

The bootstrap method argues that the distribution of $T = n^{1/2}\hat{\sigma}^{-1}\{\theta(\bar{X}) - \theta(\mu)\}$ may be approximated by the conditional distribution of $T^* = n^{1/2}\hat{\sigma}^{*-1}\{\theta(\bar{X}^*) - \theta(\bar{X})\}$, given the sample. (If $\hat{\sigma}^* = 0$, we define T^* to equal an arbitrary fixed constant c.)

Write $|x| = \{(x^1)^2 + \cdots + (x^d)^2\}^{1/2}$ for the usual Euclidean metric.

2.2. *Quantile estimation by uniform bootstrap resampling.* We shall need the following regularity conditions:

(2.1) First derivatives of θ are bounded and Hölder continuous in a neighborhood of μ,

(2.2) $$E(|X|^2) < \infty \quad \text{and} \quad \limsup_{|t| \to \infty} \left| E\{\exp(it^T X)\} \right| < 1.$$

The second part of (2.2) is Cramér's condition; see [2], page 207. Condition

(2.1) and the first part of (2.2) are sufficient to ensure that T is asymptotically normal $N(0,1)$.

Let $\mathscr{X}_b = \{X_{b1}^*, \dots, X_{bn}^*\}$, $1 \le b \le B$, be independent resamples drawn randomly, with replacement, from $\mathscr{X}$. Define

$$(2.3)\quad \bar{X}_b^* = n^{-1}\sum_{k=1}^{n} X_{bk}^*, \qquad \hat{\sigma}_b^{ij*} = n^{-1}\sum_{k=1}^{n}\left(X_{bk}^* - \bar{X}_b^*\right)^i\left(X_{bk}^* - \bar{X}_b^*\right)^j,$$

$$(2.4)\quad \hat{\sigma}_b^{*2} = \sum_{i=1}^{d}\sum_{j=1}^{d}\theta_i\left(\bar{X}_b^*\right)\theta_j\left(\bar{X}_b^*\right)\hat{\sigma}_b^{ij*},$$

$$(2.5)\quad T_b^* = n^{1/2}\hat{\sigma}_b^{*-1}\left\{\theta\left(\bar{X}_b^*\right) - \theta(\bar{X})\right\},$$

except that T_b^* is defined to equal an arbitrary fixed constant c_0 if $\hat{\sigma}_b^* = 0$. Write G_B for the empiric distribution function of $T_1^*, \dots, T_B^*$, and put

$$\xi_p = \inf\{x\colon P(T \le x) \ge p\},$$

$$(2.6)\quad \tilde{\xi}_{p,\infty} = \inf\left\{x\colon P\left(T_1^* \le x \mid \mathscr{X}\right) \ge p\right\},$$

$$\tilde{\xi}_{p,B} = \inf\{x\colon G_B(x) \ge p\}, \qquad 0 < p < 1.$$

For each fixed n, $\tilde{\xi}_{p,B} \to \tilde{\xi}_{p,\infty}$ as $B \to \infty$, with probability 1 conditional on $\mathscr{X}$. Our next result presents more detail about this last relation, as $B, n \to \infty$ together. Let Φ denote the univariate standard normal distribution function, put $\varphi = \Phi'$, and write z_p for the solution of $\Phi(z_p) = p$.

THEOREM 2.1. *Assume conditions* (2.1) *and* (2.2), *and that* $n^{\varepsilon} \le B = B(n) \le n^{\lambda}$ *for some fixed* $0 < \varepsilon < \lambda < \infty$. *Let* $0 < p < 1$. *Then*

$$(2.7)\quad \tilde{\xi}_{p,B} - \tilde{\xi}_{p,\infty} = \frac{p - G_B\left(\tilde{\xi}_{p,\infty}\right)}{\varphi(z_p) + \delta_n} + \Delta_n,$$

where $\delta_n \to 0$, $\Delta_n = O\{(B^{-1}\log B)^{3/4}\}$ *with probability* 1.

Both δ_n and Δ_n are random variables. Formula (2.7) is of use in establishing convergence properties for the sequence $\tilde{\xi}_{p,B} - \tilde{\xi}_{p,\infty}$, $n \ge 1$. For example, it is easily proved that if $n \to \infty$ and $B \to \infty$ then

$$P\left[B^{1/2}\left\{G_B\left(\tilde{\xi}_{p,\infty}\right) - p\right\} \le \{p(1-p)\}^{1/2}x \,\middle|\, \mathscr{X}\right] \to \Phi(x)$$

with probability 1, for $-\infty < x < \infty$. Therefore, assuming the conditions of Theorem 2.1, $\tilde{\xi}_{p,B} - \tilde{\xi}_{p,\infty}$ is asymptotically normally distributed with zero mean and variance $p(1-p)/B\varphi(z_p)^2$, both conditional on $\mathscr{X}$ and unconditionally.

It is readily proved via Bernstein's inequality and the Borel–Cantelli lemma that $G_B(\tilde{\xi}_{p,\infty}) - p = O\{(B^{-1}\log B)^{1/2}\}$ with probability 1. Therefore by (2.7), under the conditions of Theorem 2.1,

$$\tilde{\xi}_{p,B} - \tilde{\xi}_{p,\infty} = O\{(B^{-1}\log B)^{1/2}\}$$

with probability 1. Furthermore, if all simulations are carried out totally independently of one another, and if $B \sim \text{const}\, n^{\alpha}$, then it follows from (2.7) that with probability 1, $\pm(B/\log B)^{1/2}(\tilde{\xi}_{p,B} - \tilde{\xi}_{p,\infty})$ has lim sup equal to $\{2\alpha^{-1}p(1-p)\}^{1/2}/\varphi(z_p)$.

2.3. *Quantile estimation by importance bootstrap resampling.* We follow the prescription given by Johns [5], and refer the reader to that paper for a general discussion of importance sampling. Our description is confined to the mechanics of importance resampling.

Define $q_i = n^{-1}\exp(-\delta_i)$, $1 \le i \le n$, where

$$\delta_i = An^{-1/2}\hat{\sigma}^{-1}\sum_{j=1}^{d}\theta_j(\bar{X})(X_i - \bar{X})^j + n^{-1}C, \tag{2.8}$$

A is a fixed real number and C is chosen so that $\Sigma_i q_i = 1$. [The latter constraint entails $C = \frac{1}{2}A^2 + O(n^{-1/2})$ with probability 1 as $n \to \infty$.] Johns' [5] constant a is identical to our $-A$. As we shall see, the optimal value of A is positive.

In a change of notation from Section 2.4, let $\mathscr{X}_b^* = \{X_{b1}^*, \dots, X_{bn}^*\}$, $1 \le b \le B$, denote independent resamples drawn by resampling from $\mathscr{X}$ in a manner which gives mean weight q_i to sample value X_i:

$$P(X_{bj}^* = X_i | \mathscr{X}) = q_i.$$

Define $\bar{X}_b^*$, $\hat{\sigma}_b^{ij*}$, $\hat{\sigma}^{*2}$ and T_b^* as in (2.6)–(2.8). Let N_{bi} denote the number of times X_i appears in $\mathscr{X}_b^*$, and put

$$V_B(x) = B^{-1}\sum_{b:\,T_b^* \le x}\exp\Big(\sum_i N_{bi}\delta_i\Big),$$

$$\hat{\xi}_{p,B} = V_B^{-1}(p) = \inf\{x: V_B(x) \ge p\}.$$

Then $\hat{\xi}_{p,B}$ is the importance resampling estimate of $\tilde{\xi}_{p,\infty}$.

Consistency of $\hat{\xi}_{p,B}$ for $\tilde{\xi}_{p,\infty}$ is almost trivial to prove. To appreciate the argument, assume all the X_i's are distinct, define $\mathscr{X}_{(1)}, \dots, \mathscr{X}_{(N)}$ to be the $N = \binom{2n-1}{n}$ different resamples (ordered so that the corresponding values $t_{(j)}$ of T satisfy $t_{(1)} \le \cdots \le t_{(N)}$), write M_{ji} for the number of times X_i appears in $\mathscr{X}_{(j)}$ and put

$$\pi_j = \frac{n!}{\prod_i(M_{ji}!)}\prod_i n^{-M_{ji}}, \qquad \pi_j' = \frac{n!}{\prod_i(M_{ji}!)}\prod_i(n^{-1}e^{-\delta_i})^{M_{ji}}$$

(the probabilities associated with $t_{(j)}$ under uniform resampling, importance resampling respectively). Then $V_B \to V$ as $B \to \infty$, where

$$V(x) = \sum_{j: t_{(j)} \le x} \pi_j' \exp\Big(\sum_i M_{ji}\delta_i\Big) = \sum_{j: t_{(j)} \le x} \pi_j = G(x).$$

Therefore $\hat{\xi}_{p,B} \to G^{-1}(p) = \tilde{\xi}_{p,\infty}$. We have proved the following theorem.

THEOREM 2.2. *Assume that the distribution of X has no atoms. Then as $B \to \infty$, and conditional on $\mathscr{X}$, $\hat{\xi}_{p,B} \to \tilde{\xi}_{p,\infty}$ with probability* 1.

The definition of importance resampling which we have given is an "exact" version of an elegant "approximate" formulation which is suggested by the discussion in Johns [5]. In our context, the latter definition may be described as follows. Put

$$S_b^* = n^{1/2}\hat{\sigma}^{-1} \sum_j \theta_j(\bar{X})\big(\bar{X}_b^* - \bar{X}\big)^j,$$

and let F_B and G_B be the empiric distribution functions of $S_1^*, \dots, S_B^*$ and $T_1^*, \dots, T_B^*$, respectively. For distribution functions K, define

$$W_K(x) = \int_{-\infty}^{x} \exp(Ay + C)\, dK(y),$$

where A and C are as in (2.8). Let

$$W_K^{-1}(p) = \inf\{x : W_K(x) \ge p\}, \qquad G_B^{-1}(p) = \inf\{x : G_B(x) \ge p\},$$
$$\breve{\xi}_{p,B} = G_B^{-1}\big[F_B\big\{W_{F_B}^{-1}(p)\big\}\big].$$

Then $\breve{\xi}_{p,B}$ is an approximate form of $\hat{\xi}_{p,B}$. It converges to $G^{-1}[F\{W_F^{-1}(p)\}]$ as $B \to \infty$; this does not necessarily equal $\tilde{\xi}_{p,\infty}$.

Our next theorem describes $\hat{\xi}_{p,B} - \tilde{\xi}_{p,\infty}$ and $\breve{\xi}_{p,B} - \hat{\eta}_{p,\infty}$, where $\hat{\eta}_{p,\infty} = W_F^{-1}(p)$.

THEOREM 2.3. *Assume conditions* (2.1) *and* (2.2), *and that $n^{\varepsilon} \le B = B(n) \le n^{\lambda}$ for some fixed $0 < \varepsilon < \lambda < \infty$. Let $0 < p < 1$ and $-\infty < A < \infty$. Then we have*

(i) "*exact*" *importance resampling*:

$$\hat{\xi}_{p,B} - \tilde{\xi}_{p,\infty} = \frac{p - V_B(\tilde{\xi}_{p,\infty})}{\varphi(z_p) + \delta_n} + \Delta_{n1}; \tag{2.9}$$

(ii) "*approximate*" *importance resampling*:

$$\breve{\xi}_{p,B} - \hat{\eta}_{p,\infty} = \frac{F_B(\hat{\eta}_{p,\infty}) - G_B(\hat{\eta}_{p,\infty})}{\varphi(z_p + A) + \delta_{n1}} + \frac{p - W_{F_B}(\hat{\eta}_{p,\infty})}{\varphi(z_p) + \delta_{n2}} + \Delta_{n2}, \tag{2.10}$$

where $\delta_n, \delta_{n1}, \delta_{n2} \to 0$ *and* $\Delta_{n1}, \Delta_{n2} = O\{(B^{-1}\log B)^{3/4}\}$ *with probability* 1. *The first term on the right-hand side of* (2.10) *is negligible relative to the second, since* $B^{1/2}\{F_B(\hat{\eta}_{p,\infty}) - G_B(\hat{\eta}_{p,\infty})\} \to 0$ *in probability and* $(B/\log B)^{1/2} \cdot \{F_B(\hat{\eta}_{p,\infty}) - G_B(\hat{\eta}_{p,\infty})\} \to 0$ *with probability* 1, *as* $n \to \infty$.

Conditional on $\mathscr{X}$, $V_B(\tilde{\xi}_{p,\infty})$ and $W_{F_B}(\hat{\eta}_{p,\infty})$ are sums of independent and identically distributed random variables. By making use of this fact, we may employ formulas (2.9) and (2.10) much the same way that we did (2.7), this time to establish convergence properties of $\hat{\xi}_{p,B} - \hat{\xi}_{p,\infty}$. For example, it may be proved from the Lindeberg–Feller theorem that if $n \to \infty$ and $B \to \infty$,

$$P\Big[B^{1/2}\big\{V_B\big(\tilde{\xi}_{p,\infty}\big) - p\big\} \le \tau(A,p)x \Big| \mathscr{X}\Big] \to \Phi(x)$$

with probability 1, for $-\infty < x < \infty$, where $\tau(A,p)^2 = s^{A^2}\Phi(z_p - A) - p^2$. From this fact and Theorem 2.3 we see that $\hat{\xi}_{p,B} - \tilde{\xi}_{p,\infty}$ is asymptotically normally distributed with zero mean and variance $\tau(A,p)^2/B\varphi(z_p)^2$, both conditional on $\mathscr{X}$ and unconditionally.

Suppose $\hat{\xi}_{p,\infty} = \tilde{\xi}_{p,\infty}$ with probability 1. In Section 2.2 we showed that $\tilde{\xi}_{p,B} - \tilde{\xi}_{p,\infty}$ has asymptotic variance $p(1-p)/B\varphi(z_p)^2$, and we have just seen that $\hat{\xi}_{p,B} - \tilde{\xi}_{p,\infty}$ has asymptotic variance $\tau(A,p)^2/B\varphi(z_p)^2$. Therefore the efficiency of $\tilde{\xi}_{p,B}$ relative to $\hat{\xi}_{p,B}$ is $\rho(A,p) = \tau(A,p)^2/p(1-p)$. The minimum of $\tau(A,p)$, and hence the minimum of $\rho(A,p)$, occurs when $A = A_p$ satisfies the equation $2\Phi(z_p - A) - A^{-1}\varphi(z_p - A) = 0$. It is readily shown that this equation has exactly one solution. Table 1 lists values of A_p and $\rho(A_p, p)$ for different values of p. Now, $\rho(A_p, p)$ is increasing in p. For $p > \frac{1}{2}$, maximum efficiency in estimation of $\hat{\xi}_{p,\infty}$ is obtained by estimating the $(1-p)$th quantile of the distribution of $-T^*$ conditional on $\mathscr{X}$.

Using (2.10), we may prove that, under the conditions of Theorem 2.3, $\hat{\xi}_{p,B} - \tilde{\xi}_{p,\infty} = O\{(B^{-1}\log B)^{1/2}\}$, and that if the simulations are conducted totally independently of one another, with $B \sim \text{const}\, n^{\alpha}$, then the lim sup of $\pm(B/\log B)^{1/2}(\hat{\xi}_{p,B} - \tilde{\xi}_{p,\infty})$ equals $(2/\alpha)^{1/2}\tau(A,p)/\varphi(z_p)$, with probability 1.

TABLE 1

Values of A_p, $\rho(A_p, p)$ *for selected* p

p	A_p	$\rho(A_p, p)$
0.005	2.6561	0.0145
0.01	2.5074	0.0263
0.025	2.1787	0.0569
0.05	1.8940	0.1002
0.1	1.5751	0.1732
0.5	0.6120	0.5722
0.9	0.1150	0.8889
0.95	0.0602	0.9348
0.975	0.0320	0.9617
0.99	0.0139	0.9813
0.995	0.0074	0.9892

3. Proofs. We confine ourselves to an outline of the proof of (2.10), which is the more difficult to prove of results (2.7), (2.9) and (2.10). A key is to smooth the distributions of S^* and T^*, to eradicate their discreteness. Let ζ_b, $b \geq 1$, be independent standard normal variables, independent of everything defined so far. Let $c \geq 100$ be a fixed constant, chosen so large that $n^c \geq B(n)^{100}$ for all n. Replace $\overline{X}_b^*$ by $\overline{X}_b^o = \overline{X}_b^* + n^{-c}\zeta_b$ at each place it appears in the formulas for S_b^* and T_b^*. Let S_b^o and T_b^o denote the resulting new versions of S_b^* and T_b^*; we shall use the superscript ^{o} throughout to denote this smoothing operation. Write P', E' and Var$'$ for probability, expectation and variance, respectively, conditional on $\mathscr{X}$.

Let S^o and T^o denote generic versions of S_b^o and T_b^o; write F_B^o and G_B^o for empiric distribution functions of $S_1^o, \dots, S_B^o$ and $T_1^o, \dots, T_B^o$; and put $F^o = E'(F_B^o)$ and $G^o = E'(G_B^o)$, the distribution functions of S^o and T^o conditional on $\mathscr{X}$. Define $f^o = F^{o\prime}$ and $g^o = G^{o\prime}$, the conditional densities of S^o and T^o. Our proof is largely by a sequence of lemmas.

LEMMA 3.1. *Under the conditions of Theorem* 2.3, *and for each* $c > 0$,

$$\sup_{-\infty < x < \infty} \{|f^o(x) - \varphi(x + A)| + |f^{o\prime}(x) - \varphi'(x + A)| + |g^o(x) - \varphi(x + A)| + |g^{o\prime}(x) - \varphi'(x + A)|\} \to 0$$

with probability 1.

PROOF. Define

$$\hat{\mu} = E'\left(X_{bi}^*\right) = n^{-1} \sum_{i=1}^{n} X_i \exp(-\delta_i),$$

$$\hat{\Sigma} = \mathrm{Var}'\left(X_{bi}^*\right) = n^{-1} \sum_{i=1}^{n} (X_i - \hat{\mu})(X_i - \hat{\mu})^T \exp(-\delta_i),$$

$\hat{M} = (\hat{M}^k)$ where

$$\hat{M}^k = -A\hat{\sigma}^{-1} \sum_{j=1}^{d} \theta_j(\overline{X})\hat{\sigma}^{jk}, \qquad \hat{\Sigma}_1 = n^{-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)^T.$$

Since $\delta_1 = An^{-1/2}\hat{\sigma}^{-1}\Sigma_j\theta_j(\overline{X})(X_i - \overline{X})^j + n^{-1}C$, where C is chosen so that $\Sigma \exp(-\delta_i) = n$, then $C = \frac{1}{2}A^2 + o(1)$,

$$\hat{\mu} = \overline{X} + n^{-1/2}\hat{M} + o(n^{-1/2}), \qquad \hat{\Sigma} = \hat{\Sigma} + o(1). \tag{3.1}$$

Therefore the vector $U = n^{1/2}(\overline{X}_1^o - \overline{X})$ satisfies $E'(U) = \hat{M} + o(1)$ and $\mathrm{Var}'(U) = \hat{\Sigma}_1 + o(1)$. Write $\hat{f}$ for the density of U conditional on $\mathscr{X}$, and $\hat{f}_o$ for the normal density with mean $\hat{M}$ and variance $\hat{\Sigma}_1$. Arguing as in the first part of Appendix 2 of [3], we see that it suffices to show that

$$\sup(1 + |x|^2)\left|\hat{f}(x) - \hat{f}_o(x)\right| \to 0, \tag{3.2}$$

$$\sup(1 + |x|^2)\left|(\partial/\partial x^i)\{\hat{f}(x) - \hat{f}_o(x)\}\right| \to 0, \quad \text{each } i. \tag{3.3}$$

Observe that $n^{1/2}(X_1^* - \bar{X})$ has characteristic function, conditional on $\mathscr{X}$, given by $\psi(t) = e^{-C}\chi(t)^n$, where $\chi(t) = n^{-1}\sum_k J(t, X_k, -A\hat{\sigma}^{-1}, \bar{X})$, $J(t, x, \lambda, \mu) = \exp\{n^{-1/2}u(t, x, \lambda, \mu)\}$, $u(t, x, \lambda, \mu) = \{\lambda v(\mu) + it\}^T(x - \mu)$ and $v(\mu)$ is the d-vector whose jth element is $\theta_j(\mu)$. The characteristic function of U, conditional on $\mathscr{X}$, is $\Psi(t)\exp(-\frac{1}{2}n^{-2c}t^Tt)$. Hence by the inversion theorem for characteristic functions,

$$(3.4) \qquad \hat{f}(x) = (2\pi)^{-d}\int_{\mathbb{R}^d} e^{-C}\chi(t)^n \exp\left(-\tfrac{1}{2}n^{-2c}t^Tt - it^Tx\right) dt.$$

Write χ_o for the characteristic function correspond to $\hat{f}_o$. Given a vector $\gamma = (\gamma^1, \dots, \gamma^d)^T$ of nonnegative integers satisfying $\sum \gamma^j \le 2$, define D^γ to be the differential operator $\prod_j(\partial/\partial t^j)^{\gamma^j}$. Results (3.2) and (3.3) will follow from the inversion formulas for $\hat{f}$ and $\hat{f}_o$ if

$$(3.5) \qquad \int_{\mathbb{R}^d}(1 + |t|)\left|D^\gamma\left\{e^{-C}\chi(t)^n \exp\left(-\tfrac{1}{2}n^{-2c}t^Tt\right) - \chi_o(t)\right\}\right| dt \to 0$$

with probability 1, for all such γ's. Note that $(\partial/\partial x^i)f(x)$ is obtained by differentiating under the integral sign in (3.4); this is the reason for the term $(1 + |t|)$ in (3.5). Result (3.5) may be proved by routine, although algebraically complex, methods. □

Define $\hat{\eta}^o_{p,B} = W^{-1}_{F^o_B}(p)$, $\hat{\eta}^o_{p,\infty} = W^{-1}_{F^o}(p)$, $\hat{\xi}^o_{p,B} = G^{o-1}_B\{F^o_B(\hat{\eta}^o_{p,B})\}$, $\hat{\xi}^o_{p,\infty} = G^{o-1}\{F^o(\hat{\eta}^o_{p,\infty})\}$.

LEMMA 3.2. *Under the conditions of Theorem* 2.3,

$$\left|\hat{\eta}^o_{p,B} - \hat{\eta}^o_{p,\infty}\right| = O\left\{\left(B^{-1}\log B\right)^{1/2}\right\}$$

with probability 1 *as* $n \to \infty$.

Define $\eta_1 = \hat{\eta}^o_{p,B}$, $L(\eta) = W_{F^o}(\eta)$, $L_B(\eta) = W_{F^o_B}(\eta)$ and $D(\eta) = L_B(\eta) - L_B(\eta_1) - \{L(\eta) - L(\eta_1)\}$.

LEMMA 3.3. *Under the conditions of Theorem* 2.3 *and for each* $C_1 > 0$,

$$\sup_{|\eta - \eta_1| \le C_1(B^{-1}\log B)^{1/2}} D(\eta) = O\left\{\left(B^{-1}\log B\right)^{3/4}\right\}$$

with probability 1 *as* $n \to \infty$.

PROOF. Observe that

$$D(\eta) = B^{-1}\sum_{b=1}^{B} U_b(\eta)$$

where

$$U_b(\eta) = \int_{\eta_1}^{\eta} \exp(Ax + C)\, d\{I(S_b^o \le x) - F^o(x)\},$$

and note that $\eta_1 \to z_p$ as $n \to \infty$. If $|\eta_1 - z_p| \le C_2$ and $|\eta - z_p| \le C_2$ then conditional on χ, the random variables $U_b(\eta)$ are independent and identically distributed with

$$|U_b(\eta)| \le C_3, \qquad E'\{U_b(\eta)\} = 0, \qquad \mathrm{Var}'\{U_b(\eta)\} \le C_4|\eta - \eta_1|.$$

Application of Bernstein's inequality ([4], page 17) now gives

$$\sup_{|\eta - \eta_1| \le C_1(B^{-1}\log B)^{1/2}} P'\left\{|D(\eta)| > C_6(B^{-1}\log B)^{3/4}\right\}$$

$$\le 2\exp(-C_5 C_6^2 \log B),$$

for n sufficiently large.

Let $\eta^{(1)}, \eta^{(2)}, \dots$ be a sequence such that $\eta_0 - C_2 = \eta^{(1)} < \eta^{(2)} < \cdots < \eta^{(B_1 - 1)} < \eta_0 + C_2 \le \eta^{(B_1)}$ and $\eta^{(i+1)} - \eta^{(i)} = B^{-1}$ for $i \ge 1$. Then $B_1 \sim 2C_2 B$. Given $\xi \in (\eta_0 - C_2, \eta_0 + C_2)$, let i_ξ denote that integer between 1 and B_1 such that $\eta^{(i_\xi - 1)} < \xi \le \eta^{(i_\xi)}$. It may be proved from Lemma 3.1 and Bernstein's inequality that

$$\sup_{|\xi - z_p| \le C_2} |D(\eta^{(i_\xi)}) - D(\xi)| = O(B^{-1}\log B),$$

$$\sup_{i: |\eta^{(i)} - \eta_1| \le C_1(B^{-1}\log B)^{1/2}} |D(\eta^{(i)})| = O\left\{(B^{-1}\log B)^{3/4}\right\}$$

almost surely. The lemma follows from these two results. □

Take $\eta_1 = \hat{\eta}^o_{p,\infty}$, and fix $C_1 > 0$. By Lemma 3.1, L' and L'' are uniformly bounded on any finite set. Furthermore, $\eta_1 \to z_p$. Hence

$$|L(\eta) - L(\eta_1) - (\eta - \eta_1)L'(\eta_1)| \le A_n(\eta - \eta_1)^2$$

uniformly in $|\eta - \eta_1| \le C_1$, where $A_n > 0$ denotes a random variable satisfying

$$\limsup_{n \to \infty} A_n < \infty$$

almost surely. Take $\eta = \hat{\eta}^o_{p,B}$. Then by Lemma 3.2,

$$\eta - \eta_1 = O\left\{(B^{-1}\log B)^{1/2}\right\},$$

and so by Lemma 3.3,

$$L(\eta) - L(\eta_1) = L_B(\eta) - L_B(\eta_1) + O\left\{(B^{-1}\log B)^{3/4}\right\}.$$

But $L_B(\eta) = p + O(B^{-1})$, and $L'(\eta_1)$ is bounded away from 0 (by Lemma 3.1), so

$$\eta - \eta_1 = \frac{p - L_B(\eta_1)}{L'(\eta_1)} + O\Big\{\big(B^{-1}\log B\big)^{3/4}\Big\}.$$

Define $\eta_1 = \hat{\eta}^o_{p,\infty}$, $\xi_1 = \hat{\xi}^o_{p,\infty} = G_o^{-1}\{F^o(\hat{\eta}^o_{p,\infty})\}$,

$$D_1(\eta) = F^o_B(\eta) - F^o_B(\eta_1) - \{F^o(\eta) - F^o(\eta_1)\},$$

$$D_2(\xi) = G^o_B(\xi) - G^o_B(\xi_1) - \{G^o(\xi) - G^o(\xi_1)\},$$

Our next two lemmas may be proved much as was Lemma 3.3.

LEMMA 3.4. *Under the conditions of Theorem* 2.3 *and for each* $C_1 > 0$,

$$\sup_{|\eta-\eta_1|\le C_1(B^{-1}\log B)^{1/2}} |D_1(\eta)| = O\Big\{\big(B^{-1}\log B\big)^{3/4}\Big\},$$

$$\sup_{|\xi-\xi_1|\le C_1(B^{-1}\log B)^{1/2}} |D_2(\xi)| = O\Big\{\big(B^{-1}\log B\big)^{3/4}\Big\},$$

with probability 1 *as* $n \to \infty$.

LEMMA 3.5. *Under the conditions of Theorem* 2.3 *and for each* $C_1 > 0$,

$$\sup_{|x|\le C_1} \{|F^o_B(x) - F^o(x)| + |G^o_B(x) - G^o(x)|\} = O\Big\{\big(B^{-1}\log B\big)^{1/2}\Big\}$$

with probability 1 *as* $n \to \infty$.

Lemmas 3.1, 3.2, 3.4 and 3.5 may be combined to prove Lemma 3.6.

LEMMA 3.6. *Under the conditions of Theorem* 2.3,

$$F^o_B(\hat{\eta}^o_{p,B}) - F^o(\hat{\eta}^o_{p,\infty}) = O\Big\{\big(B^{-1}\log B\big)^{1/2}\Big\},$$

$$F^o_B(\hat{\eta}^o_{p,B}) - F^o(\hat{\eta}^o_{p,\infty}) = \frac{\{p - L_B(\hat{\eta}^o_{p,\infty})\}f^o(\hat{\eta}^o_{p,\infty})}{L'(\hat{\eta}^o_{p,\infty})} + O\Big\{\big(B^{-1}\log B\big)^{3/4}\Big\} \tag{3.6}$$

with probability 1 *as* $n \to \infty$.

Our next lemma is a straightforward consequence of Lemmas 3.1, 3.5 and 3.6.

LEMMA 3.7. *Under the conditions of Theorem* 2.3,

$$\hat{\xi}^o_{p,B} - \hat{\xi}^o_{p,\infty} = O\Big\{\big(B^{-1}\log B\big)^{1/2}\Big\}$$

with probability 1 *as* $n \to \infty$.

Define $\xi_1 = G_o^{-1}\{F^o(\hat{\eta}^o_{p,\infty})\}$, $\xi = \hat{\xi}^o_{p,B}$. By Lemmas 3.1, and 3.4 and 3.7,

$$\xi - \xi_1 = \frac{F^o_B(\hat{\eta}^o_{p,B}) - G^o_B(\xi_1)}{g^o(\xi_1)} + O\left\{\left(B^{-1}\log B\right)^{3/4}\right\}.$$

From this result, (3.6) and the fact that $L'(x) = \exp(Ax + C)g^o(x)$, we conclude that

$$\hat{\xi}^o_{p,B} - \hat{\xi}^o_{p,B} = \frac{F^o_B(\hat{\eta}^o_{p,\infty}) - G^o_B\left(\hat{\xi}^o_{p,\infty}\right)}{g^o\left(\hat{\xi}^o_{p,\infty}\right)} + \frac{p - L_B(\hat{\eta}^o_{p,\infty})}{g^o\left(\hat{\xi}^o_{p,\infty}\right)\exp\left(A\hat{\eta}^o_{p,\infty} + C\right)} + O\left\{\left(B^{-1}\log B\right)^{3/4}\right\}$$

with probability 1. Using Lemma 3.1 to simplify the denominators on the right-hand side and noting that $\hat{\xi}^o_{p,\infty}$ and $\hat{\eta}^o_{p,\infty}$ both converge to z_p with probability 1, we conclude that

$$\hat{\xi}^o_{p,B} - \hat{\xi}^o_{p,\infty} = \frac{F^o_B(\hat{\eta}^o_{p,\infty}) - G^o_B\left(\hat{\xi}^o_{p,\infty}\right)}{\varphi(z_p + A) + \delta^o_{n1}} + \frac{p - L_B(\hat{\eta}^o_{p,\infty})}{\varphi(z_p) + \delta^o_{n2}} + O\left\{\left(B^{-1}\log B\right)^{3/4}\right\},$$

where $\delta^o_{n1} \to 0$ and $\delta^o_{n2} \to 0$ with probability 1. This leads readily to (2.10). □

Acknowledgments. I am grateful to M. Vernon Johns for providing a copy of his paper [5] prior to publication, to Alan H. Welsh for several helpful conversations on the subject of Bahadur representations and to two referees and an associate editor for their particularly constructive comments. My thanks go also to Daryl J. Daley for editorial assistance.

REFERENCES

[1] BAHADUR, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577–580.

[2] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

[3] HALL, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.* **14** 1453–1462.

[4] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[5] JOHNS, M. V. (1988). Importance sampling for bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **83** 709–714.

[6] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA ACT 2601
AUSTRALIA