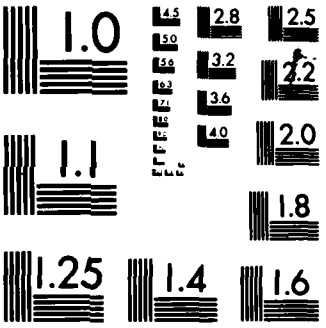
AD-A158 880 ON THE NUMBER OF BOOTSTRAP SIMULATIONS REQUIRED TO CONSTRUCT A CONFIDENCE (U) NORTH CAROLINA UNIV AT CHAPEL HILL CENTER FOR STOCHASTIC PROC P HALL MAR 85 1/1

UNCLASSIFIED TR-97 AFOSR-TR-85-0678 F49620-82-C-0009 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

(2)

RT DOCUMENTATION PAGE

AD-A158 880

1a.	1b. RESTRICTIVE MARKINGS
2a.	3. DISTRIBUTION/AVAILABILITY OF REPORT Unlimited [illegible] release; distribution unlimited.
2b.	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report No. 97	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 85-0678
6a. NAME OF PERFORMING ORGANIZATION Center for Stochastic Processes 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code) Dept. of Statistics University of North Carolina Chapel Hill, NC 27514	7b. ADDRESS (City, State and ZIP Code) Bolling Air Force Base Washington, DC 20332 SEP 10 1985
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-82-C-0009.
8c. ADDRESS (City, State and ZIP Code) Bolling Air Force Base Washington, DC 20332	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 61102F PROJECT NO.: 2304 TASK NO.: A5 WORK UNIT NO.:
11. TITLE (Include Security Classification) "On the number of bootstrap simulations required to construct a confidence interval"	
12. PERSONAL AUTHOR(S) P. Hall	
13a. TYPE OF REPORT technical 13b. TIME COVERED FROM 9/84 TO 8/85	14. DATE OF REPORT (Yr., Mo., Day) March 1985 15. PAGE COUNT 13
16. SUPPLEMENTARY NOTATION	
17. COSATI CODES FIELD / GROUP / SUB. GR.	18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Keywords: Bootstrap, confidence interval, number of simulations.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

We make two points about the number, B, of bootstrap simulations needed to construct a percentile-t confidence interval based on an N-sample from a continuous distribution: (i) The bootstrap's reduction of error of coverage probability, from $O(n^{-1/2})$ to $O(n^{-1})$, is available uniformly in B, provided nominal coverage probability is a multiple of $(B+1)^{-1}$. In fact, this improvement is available even if the number of simulations is held fixed as n increases. (ii) In a large sample, the simulated statistic values behave like random observations from a continuous distribution, unless B increases faster than any power of sample size. Only if B increases exponentially quickly is there a detectable effect due to discreteness of the bootstrap statistic.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Brian Woodruff	22b. TELEPHONE NUMBER (Include Area Code) (202) 767-5027 22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

AFOSR-TR- 78

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

ON THE NUMBER OF BOOTSTRAP SIMULATIONS REQUIRED TO CONSTRUCT A CONFIDENCE INTERVAL

by

Peter Hall

Technical Report No. 97

March 1985

85

ON THE NUMBER OF BOOTSTRAP SIMULATIONS REQUIRED TO CONSTRUCT A CONFIDENCE INTERVAL

by

Peter Hall[1]

University of North Carolina, Chapel Hill[2]

Summary. We make two points about the number, B, of bootstrap simulations needed to construct a percentile-t confidence interval based on an n-sample from a continuous distribution: (i) The bootstrap's reduction of error of coverage probability, from $O(n^{-1/2})$ to $O(n^{-1})$, is available uniformly in B, provided nominal coverage probability is a multiple of $(B+1)^{-1}$. In fact, this improvement is available even if the number of simulations is held fixed as n increases. (ii) In a large sample, the simulated statistic values behave like random observations from a continuous distribution, unless B increases faster than any power of sample size. Only if B increases exponentially quickly is there a detectable effect due to discreteness of the bootstrap statistic.

Key words: Bootstrap, confidence interval, number of simulations.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE
Chief, Technical Information Division

[1]Research supported by AFOSR Grant No. F 49620 82 C 0009.

[2]On leave from Australian National University.

DTIC COPY INSPECTED
A-1

1. Introduction

The purpose of this ~~note~~ paper is to make two points about the effect of the number of bootstrap simulations, B, on percentile-t bootstrap confidence intervals. The first point concerns coverage probability; the second, distance of the "simulated" critical point from the "true" critical point derived with $B=\infty$. In both cases ~~we have~~ we have in mind applications to "smooth" statistics, such as the Studentized mean of a sample drawn from a continuous distribution. ~~We shall~~ We indicate the changes that have to be made if the distribution of the statistic is not smooth.

To make our point about coverage probability, recall that if we conduct B bootstrap simulations, the resulting statistic values divide the real line into B+1 parts. Therefore in principle, confidence intervals whose critical points are based on B simulations have coverage probabilities close to nominal levels $b/(B+1)$, for $b=1,\ldots,B$. If the sample size is n and $B=\infty$, then the "Edgeworth inversion" effect of the bootstrap argument means that true coverage probability of a confidence interval whose desired coverage is α, is actually $\alpha + \delta_n(\alpha)$, and $\delta_n \equiv \sup_\alpha |\delta_n(\alpha)| = O(n^{-1})$ (Hall 1984). This is a notable improvement over the level $\alpha + O(n^{-1/2})$ offered by traditional methods. Strikingly, this improvement is available for *any* value of B, *even for fixed* B. In fact, if δ_n^* is the worst possible error between true coverage probability and nominal coverage probability when only B simulations are used, then $\delta_n^* \le \delta_n$ *uniformly in* B. Therefore *the worst departure of true coverage probability from nominal coverage probability using any finite number of simulations, does not exceed the worst departure using an infinite number of simulations.*

For example, suppose we wish to construct a one-sided 90% confidence interval. The smallest value of B we can use is B=9: notice that 90% =9/(9+1).

The endpoint of the interval would be based on either the smallest or largest of the 9 simulations, depending on whether the interval was left-handed or right-handed. If we let $n \to \infty$, but always did only 9 simulations, then the coverage probability of our interval would still be $0.9 + O(n^{-1})$. So even with a fixed number of simulations, we improve on the traditional coverage probability of $0.9 + O(n^{-1/2})$.

This property forms a convenient safety-net for the bootstrap algorithm: if a statistician cannot conduct as many simulations as he would like, he can be sure that he pays virtually no penalty in terms of accuracy of coverage probability. The only penalty is in length of confidence interval - if B is small then the true critical point may stray from its limiting value when $B=\infty$, so that there will be some tendency for confidence intervals to be over-long.

In addition, B does not have to be particularly large before exact coverage probability agrees with the theoretical limit as $B \to \infty$. For example, if B equals sample size then the probabilities only disagree at the level $O(n^{-2})$.

We shall investigate these properties in Section 2. As part of our study we shall give an explicit formula for the second-order term in an expansion of coverage probability for the case of Studentized mean. That formula makes it clear that if the sample is actually normally distributed, then *even using a fixed value of* B, the coverage probability of a bootstrap confidence interval for population mean differs from the nominal level by only $O(n^{-2})$.

We shall also investigate the effect of the size of B on critical points. Remember that the distribution of the bootstrap statistic is discrete. Beran (1984), among others, has pondered the use of smoothing techniques to overcome discreteness. In Section 3 we shall show that under a very weak smoothness assumption, even weaker than continuity, the distribution of the simulated

bootstrap statistic behaves like a continuous distribution with a density uniformly close to the standard normal density. In fact, the error in this continuous approximation to the discrete bootstrap distribution is of order $n^{-\lambda}$ for all $\lambda>0$. We shall show that the number of bootstrap simulations, B, has to be an exponentially large function of sample size before the discreteness of the bootstrap distribution becomes apparent.

On the other hand, if the sampling distribution is lattice then it is easily seen that the atoms of the bootstrap statistic are of order $n^{-1/2}$, and then it is essential to smooth the distribution of the bootstrap statistic. Our results on coverage probability have analogues for lattice-valued statistics, but it should be remembered that in that case, rounding error reduces approximation order from n^{-1} to only $n^{-1/2}$.

We shall confine attention to one-sided, percentile-t confidence intervals. The bias-corrected percentile technique (Efron 1979) is suitable only for two-sided intervals, and in that situation traditional methods, percentile methods and bias-corrected percentile methods all give coverage probabilities whose errors are of order n^{-1}. There, the advantages of the bootstrap cannot be reported so clearly in terms of coverage probability.

Related work includes that of Beran (1982,1984), Singh (1981) and Babu and Singh (1983), who have studied large-sample properties of the bootstrap algorithm. The latter two papers are concerned with conditional Edgeworth expansions. In Section 2 we shall briefly mention Edgeworth expansions, but those considered here are *unconditional*.

2. Coverage probability

Let $\hat{\theta} = \hat{\theta}(X_1,\ldots,X_n)$ be a $\sqrt{n}$-consistent estimator of a parameter θ, based on a random sample $X = \{X_1,\ldots X_n\}$. Let $n^{-1}\hat{\sigma}^2(X_1,\ldots,X_n)$ be a consistent estimator of the variance of $\hat{\theta}$. We shall consider confidence intervals for θ based on the statistic $T = n^{1/2}(\hat{\theta}-\theta)/\hat{\sigma}$.

Let $Y_1,\ldots,Y_n$ be independent and identically distributed, conditional on X, with distribution $P(Y = X_j|X) = n^{-1}$, $1 \le j \le n$. The bootstrap statistic is obtained by using the sample $Y \equiv \{Y_1,\ldots,Y_n\}$ in place of X. Thus, we consider $\hat{\theta}^* \equiv \hat{\theta}(Y_1,\ldots,Y_n)$, $\hat{\sigma}^* \equiv \hat{\sigma}(Y,\ldots,Y_n)$, $T^* \equiv n^{1/2}(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$. We may work out the distribution of T^*, conditional on X, to arbitrary accuracy by means of simulation. Thus, we may define

$$t_\alpha \equiv \inf\{t: P(T^* \le t|X) \ge \alpha\},$$

which is the bootstrap approximation to that point x_α such that $P(T \le x_\alpha) = \alpha$. For example, the "optimal" but unattainable interval $I_0 = [\hat{\theta} - n^{-1/2}x_\alpha\hat{\sigma},\infty)$ covers θ with probability α; the interval $I_1 \equiv [\hat{\theta} - n^{-1/2}t_\alpha\hat{\sigma},\infty)$ covers θ with probability $\alpha+O(n^{-1})$.

In practice the value of t_α is usually itself estimated, by simulation. Conditional on X, let $T_1^*,\ldots,T_B^*$ be independent copies of T^*. Arrange them in ascending order: $T^*_{(1)} \le \cdots \le T^*_{(B)}$. Suppose we select $T^*_{(\nu+1)}$ as our approximation to t_α, for a given integer $0 \le \nu \le B-1$. Let $p \equiv P(T^* < T|X)$, and conditional on X, let N have the binomial $Bi(B,p)$ distribution. In place of I_1, we would use the interval $I_2 \equiv [\hat{\theta} - n^{-1/2}T^*_{(\nu+1)}\hat{\sigma},\infty)$. Conditional on X, the chance that I_2 covers θ is

$$P(T \le T^*_{(\nu+1)}|X) = P(\text{at most } \nu \text{ out of } T_1^*,\ldots,T_B^* \text{ are} < T|X)$$

$$= P(N \le \nu|X) = \sum_{j=0}^{\nu}\binom{B}{j}p^j(1-p)^{B-j}.$$

Therefore the exact, unconditional coverage probability of I_2, is

$$a(\nu,B) \equiv \sum_{j=0}^{\nu} \binom{B}{j} \int_0^1 u^j(1-u)^{B-j} dP(p \le u). \tag{2.1}$$

Had we been able to do an indefinite amount of simulation , we would have taken $\nu \sim \alpha B$ as $B \to \infty$, and obtained the interval I_1 whose coverage probability is

$$\lim_{B\to\infty} a(\nu,B) = P(p \le \alpha).$$

Our aim is to determine how close $a(\nu,B)$ is to its limit, $P(p \le \alpha)$.

Hall (1984) has described expansions for the distribution of p in general circumstances. Those results show that p has asymptotically a uniform distribution, and that

$$P(p \le \alpha) = \alpha + n^{-1}R_n(\alpha) \tag{2.2}$$

where R_n is bounded uniformly in $n \ge 1$ and $0 < \alpha < 1$. Define

$$G(u) = \sum_{j=0}^{\nu} \binom{B}{j} u^j(1-u)^{B-j}.$$

Then by (2.1),

$$a(\nu,B) = (\nu+1)(B+1)^{-1} + n^{-1}\int_0^1 G(u)\, dR_n(u). \tag{2.3}$$

We call $(\nu+1)(B+1)^{-1}$ the *nominal* coverage probability of confidence interval I_2. To simplify the integral in (2.3), let $\tau = \tau(\nu,v)$ be the solution of $G(\tau) = v$, for $0 < v < 1$. Then

$$\int_0^1 G(u)dR_n(u) = \int_0^1 dR_n(u) \int_0^{G(u)} dv = \int_0^1 dv \int_0^{\tau} dR_n(u) = \int_0^1 R_n(\tau)dv,$$

and so

$$a(\nu,B) = (\nu+1)(B+1)^{-1} + n^{-1} \int_0^1 R_n(\tau)dv. \tag{2.4}$$

It is clear from (2.4) that if we seek a confidence interval whose coverage probability is a multiple of $(B+1)^{-1}$, then the worst error we commit if we simulate only B times is no more than $n^{-1} \sup_{\alpha} |R_n(\alpha)|$. In view of (2.2), this is the worst error committed if we simulate an *infinite* number of times. In this sense there is nothing to be gained, in terms of coverage probability, by simulating often. If we are after a 95% interval, and are not overly concerned about interval length, we might simulate B=19 times and take ν=18.

To investigate this phenomenon a little more deeply, we shall examine an asymptotic formula for $R_n(\alpha)$. Here it is convenient to concentrate on a special case, such as Studentized mean. There, $\mathcal{X} = \{X_1,\ldots,X_n\}$ is a scalar random sample from a distribution with mean μ and variance σ^2, and $\theta = \mu$, $\hat{\theta} = \bar{X} = n^{-1}\Sigma X_i$, $\hat{\sigma}^2 = n^{-1}\Sigma(X_i - \bar{X})^2$. The bootstrap argument may be used to set confidence intervals for μ without knowing σ^2. Techniques developed in Hall (1984), although now requiring much more tedious algebra , give us

$$R_n(\alpha) = \psi_1(z_\alpha)\phi(z_\alpha) + n^{-1/2}\psi_2(z_\alpha)\phi(z_\alpha) + O(n^{-1})$$

uniformly in α, where Φ is the standard normal distribution function, $\phi = \Phi'$, z_α is the solution of $\Phi(z_\alpha) = \alpha$,

$$\psi_1(z) = \frac{1}{6}\left(\frac{3}{2}\lambda_3^2 - \lambda_4\right)z(1 + 2z^2),$$

$$\begin{aligned}\psi_2(z) = &-\lambda_3 \frac{1}{48}(1 + 2z^2)(z^4 - 18z^2 - 39)\\ &- \lambda_3^3 \left\{\frac{1}{96}(1 + 2z^2)(z^4 + 4z^2 - 23) + \frac{5}{24}z^2(4z^2-1)\right\}\\ &+ \lambda_3\lambda_4 \left\{\frac{1}{288}(1+2z^2)(5z^4 + 2z^2 - 139) + \frac{z^2}{36}(36z^2 - 23)\right\}\\ &- \lambda_5 \left\{\frac{1}{144}(1+2z^2)(z^4 - 6z^2 - 33) + \frac{1}{12}z^2(z^2 - 3)\right\},\end{aligned}$$

and λ_j is the standardized j'th cumulant; for example, $\lambda_3 = E(X-\mu)^3\sigma^{-3}$ and $\lambda_4 = E(X-\mu)^4\sigma^{-4}-3$. An outline of the argument is given in Appendix (i). For simplicity, define $Q_i(\alpha) \equiv \psi_i(z_\alpha)\phi(z_\alpha)$; then

$$P(p \leq \alpha) = \alpha + n^{-1}Q_1(\alpha) + n^{-3/2}Q_2(\alpha) + O(n^{-2}) \tag{2.5}$$

uniformly in α. From (2.4) we obtain:

$$a(\nu,B) = (\nu+1)(B+1)^{-1} + n^{-1}\int_0^1 Q_1(\tau)dv + n^{-3/2}\int_0^1 Q_2(\tau)dv + O(n^{-2}) \tag{2.6}$$

uniformly in $0 \leq \nu \leq B-1$ and $B \geq 1$. A little asymptotic analysis based on the normal approximation to the binomial shows that

$$\tau(\nu,v) = \beta-B^{-1/2}\{\beta(1-\beta)\}^{1/2}z_v + B^{-1}\frac{1}{6}(1-2\beta)(1+2z_v^2) + o(B^{-1}),$$

where $\beta = (\nu + \frac{1}{2})B^{-1}$. This expansion holds uniformly in values $v \in (B^{-2},1-B^{-2})$. Substituting into (2.6) and noting that $\int z_v dv = 0$, we get:

$$a(\nu,B) = \alpha' + n^{-1}Q_1(\alpha') + n^{-3/2}Q_2(\alpha') + O(n^{-1}B^{-1} + n^{-2}),$$

where $\alpha' = (\nu+1)(B+1)^{-1}$. This expansion is virtually identical to (2.5).

If B is chosen so that nominal coverage probability equals α, then $a(\nu,B)$ and $P(p \leq \alpha)$ agree to <u>second</u> order if B is of larger order than square root of sample size, and to <u>third</u> order if B is of larger order than sample size.

3. Critical Point

For the sake of definiteness we shall concentrate on the Studentized mean. We shall impose Cramér's smoothness condition on the pair (X,X^2):

$$\limsup_{|s|+|t|\to\infty} |E \exp(isX + itX^2)| < 1. \tag{3.1}$$

This condition holds for any random variable X whose distribution has a nontrivial continuous component, and also for certain singular distributions.

Our initial aim is to investigate the smoothness of the distribution of T*, conditional on $\mathcal{X}$. Of course, T* has a discrete distribution, with atoms determined by the sample $\mathcal{X}$. We may artificially smooth that distribution by adding small, continuous errors to the simulated sample points Y_i. For example, take $\ell > 0$ and let $N_1,\ldots,N_n$ be independent $N(0,n^{-2\ell})$ random variables independent of $\mathcal{X}$ and $\mathcal{Y}$. (Remember: $\mathcal{X}= \{X_1,\ldots,X_n\}$, $\mathcal{Y}= \{Y_1,\ldots,Y_n\}$.) Set $Z_i \equiv Y_i+N_i, \bar{Z} \equiv n^{-1}\Sigma Z_i$ and

$$T' \equiv n^{1/2}(\bar{Z} - \bar{X})/(n^{-1}\Sigma Z_i^2 - \bar{Z}^2)^{1/2}.$$

The presence of the smooth perturbations N_i means that conditional on $\mathcal{X}$, T' has a continuous distribution with density g, say. Given any $\lambda > 0$ we may choose ℓ so large that with probability one,

$$P(|T' - T^*| > n^{-\lambda}|\mathcal{X}) = O(n^{-\lambda}) \tag{3.2}$$

as $n\to\infty$. In this sense, the *discrete* random variable T* may be approximated by a *continuous* variable T', with an error of order $n^{-\lambda}$ for arbitrarily large λ.

At first sight this approximation seems spurious, and the reader is justified in being very skeptical. It seems likely that the density of g will closely track the atoms of the discrete distribution of T*, and so be quite unsmooth. After all, the continuous approximation is only supported by minute perturbations N_i, which are shrinking to zero at a rate of $n^{-\ell}$ for arbitrarily large ℓ.

However, the theorem below shows that the density g is actually quite smooth. In fact, *no matter how large the value of* ℓ, g uniformly approximates the standard normal density ϕ.

Theorem 3.1. *Assume condition* (3.1), *and that* $E(|X|^{4+\varepsilon}) < \infty$ *for some* $\varepsilon > 0$. *Then for each* $\ell > 0$,

$$\sup_{-\infty<x<\infty} |g(x) - \phi(x)| \to 0 \tag{3.3}$$

almost surely as $n\to\infty$.

The proof uses standard techniques of Fourier inversion, and will be outlined in Appendix (ii). The key to this result is the fact that the order of the approximation in (3.2) is not required to be exponentially small. There exist constants c_n decreasing *very* rapidly to zero such that, if (3.2) holds for a continuous variable T' and with c_n replacing $n^{-\lambda}$, then the approximation at (3.3) breaks down. In that case the density g does track the atoms of T* too closely.

Theorem 3.1 implies that the simulated bootstrap values behave like values from a continuous distribution, provided B is not exponentially large. For example, suppose we conduct B simulations and use $T^*_{(\nu+1)}$ as our approximation to the true critical point t_α. Assume ν is chosen so that $\nu = \alpha B + o(B^{1/2})$ as $B\to\infty$; this is quite reasonable, since we would usually have $\nu = \alpha B + O(1)$. We shall prove below that if B increases no faster than n^λ, for any $\lambda > 0$, then as B and $n\to\infty$ the conditional probability $P\{B^{1/2}(T^*_{(\nu+1)} - t_\alpha) \le x|\mathcal{X}\}$ converges to the probability that a normal variable with zero mean and variance $\sigma^2 = \alpha(1-\alpha)/\phi^2(z_\alpha)$, does not exceed x. Therefore $B^{1/2}(T^*_{(\nu+1)} - t_\alpha)$ has a limiting $N(0,\sigma^2)$ density, conditional on $\mathcal{X}$ and also unconditionally. This is the limiting distribution of the α'th quantile from a *continuous*

distribution whose density converges uniformly to the standard normal density. The result will fail if B increases too quickly, but B *has to increase faster than any power of sample size before the discreteness of the distribution of* T* *becomes apparent.* Therefore, provided the sampling distribution is continuous, we seldom need to smooth before constructing critical points.

To prove the result stated in the previous paragraph, let $y \equiv t_\alpha + B^{-1/2}x$ and $q \equiv P(T^* \le y|\mathcal{X})$, and observe that with probability one,

$$P\{B^{1/2}(T^*_{(\nu+1)} - t_\alpha) \le x|\mathcal{X}\} = P(T^*_{(\nu+1)} \le y|\mathcal{X})$$

$$= \sum_{j=\nu+1}^{B} \binom{B}{j} q^j (1-q)^{B-j} = 1 - \Phi[(\nu+1-Bq)\{Bq(1-q)\}^{-1/2}] + o(1), \qquad (3.4)$$

using the normal approximation to the binomial. If we show that with probability one,

$$q = \alpha + B^{-1/2}x\phi(z_\alpha) + o(B^{-1/2}), \qquad (3.5)$$

then it will follow that the right-hand side of (3.4) converges almost surely to $\Phi[x\phi(z_\alpha)\{\alpha(1-\alpha)\}^{-1/2}]$, as required. Choose $\lambda > 0$ so large that $B/n^\lambda \to 0$, and let T' be as in (3.2). In view of (3.2), result (3.5) will follow if we show that for each $-\infty < x < \infty$, and with probability one,

$$P(T' \le t_\alpha + B^{-1/2}x|\mathcal{X}) = \alpha + B^{-1/2}x\phi(z_\alpha) + o(B^{-1/2}). \qquad (3.6)$$

But Theorem 3.1 implies that

$$P(T' \le t_\alpha + B^{-1/2}x|\mathcal{X}) = P(T' \le t_\alpha|\mathcal{X}) + B^{-1/2}x\phi(t_\alpha) + o(B^{-1/2}). \qquad (3.7)$$

It also follows from the theorem, and from the definition of t_α, that

$$P(T' \le t_\alpha|\mathcal{X}) \le P(T' \le t_\alpha - 2B^{-1}|\mathcal{X}) + O(B^{-1})$$

$$\le P(T \le t_\alpha - B^{-1}|\mathcal{X}) + o(B^{-1/2}) \le \alpha + o(B^{-1/2}),$$

and likewise $P(T' \leq t_\alpha | \mathcal{X}) \geq \alpha + o(B^{-1/2})$. Therefore $P(T' \leq t_\alpha | \mathcal{X}) = \alpha + o(B^{-1/2})$. Result (3.6) is now immediate from (3.7).

Appendix (i): Verification of (2.5)[1]

The proof is similar to Hall (1984), although with considerably greater complexity. The only smoothness condition required is (3.1). To identify functions Q_1 and Q_2, let T be the Studentized mean and let π_1, π_2, π_3 be polynomials defined by the following inverse Cornish-Fisher expansion:

$$P\{T \leq x + n^{-1/2}\pi_1(x) + n^{-1}\pi_2(x) + n^{-3/2}\pi_3(x)\} = \Phi(x) + O(n^{-2}).$$

Formulae may be derived using results of Geary (1947); for example,

$$\pi_2(x) = z\{(5/72)\lambda_3^2\,(4z^2-1) - (1/12)\lambda_4(z^2-3) + (1/4)(z^2+1)\}.$$

Let Π_j denote the version of π_j in which λ_j is replaced by its sample estimate $\hat{\lambda}_j$; for example, $\hat{\lambda}_4 = \hat{\sigma}^{-4}n^{-1}\,\Sigma(X_i-\overline{X})^4 - 3$. The functions Q_1, Q_2 are obtainable from the relation

$$\begin{aligned} P\{T \leq z_\alpha + n^{-1/2}\Pi_1(z_\alpha) + n^{-1}\Pi_2(z_\alpha) + n^{-3/2}\Pi_3(z_\alpha)\} &= P(p \leq \alpha) + O(n^{-2}) \\ &= \alpha + n^{-1}Q_1(\alpha) + n^{-3/2}Q_2(\alpha) + O(n^{-2}). \end{aligned} \tag{A.1}$$

First find the cumulants of the random variable

$$S(\alpha) \equiv T - n^{-1/2}\Pi_1(z_\alpha) - n^{-1}\Pi_2(z_\alpha) - n^{-3/2}\Pi_3(z_\alpha)$$

to order n^{-2}; then use the cumulants to obtain an Edgeworth expansion of $P\{S(\alpha) \leq x\}$ to order n^{-2}; and finally set $x = z_\alpha$, to obtain formula (2.5) via (A.1).

(1) For easy comparison with classical literature, we assume sample variance has divisor n-1, not n. Notice that (2.5) is invariant under changes of scale of T.

Appendix (ii): Proof of Theorem 3.1.

Without loss of generality, $E(X)=0$. Let $W_i = Z_i-\overline{X}$, $w^2=E(W_1^2|\mathcal{X})$, $\underset{\sim}{U}_i = (W_i, W_1^2-w^2)$, $\underset{\sim}{U} = n^{-1/2}\Sigma\underset{\sim}{U}_i$, $\underset{\sim}{\Sigma} = \mathrm{var}(\underset{\sim}{U}_1|\mathcal{X})$, f be the density of $\underset{\sim}{U}$ conditional on $\mathcal{X}$, f_0 be the conditional density of the bivariate normal distribution with zero mean and covariance $\underset{\sim}{\Sigma}$, and χ,χ_0 be the characteristic functions of f,f_0, respectively. Notice that

$$g(x)=\int_{n^{-1/2}}^{\infty}\{(1+n^{-1/2}v)(1+n^{-1}x^2)^{-3}\}^{1/2}\, w^3 f\{wu(x,v),w^2v\}dv, \tag{A.2}$$

where $u(x,v) \equiv x\{(1+n^{-1/2}v)(1+n^{-1}x^2)^{-1}\}^{1/2}$. Define g_0 by (A.2) but with f_0 replacing f. It is easily proved that $\sup|g_0-\phi| \overset{a.s.}{\to} 0$, and so it suffices to show $\sup|g-g_0| \overset{a.s.}{\to} 0$. For this, we may show

$$\sup\,(1+|y|^2)|f(y)-f_0(y)| \overset{a.s.}{\to} 0.$$

That result follows by Fourier inversion if we prove that for nonnegative integer vectors $\gamma = (\gamma_1,\gamma_2)$ with $\gamma_1+\gamma_2 \le 2$, $\int|D^\gamma(\chi-\chi_0)| \overset{a.s.}{\to} 0$, where D is the differential operator. We treat only $\gamma = 0$; other cases are similar.

Characteristic function manipulations common to estimates of rates of convergence show that for some small $\eta > 0$, and for all sufficiently large n,

$$\sup_{|t|\le\eta n^{1/2}} |\chi_n(t)-\chi_0(t)|e^{\eta t^2} \le \eta^{-1}n^{-\eta}$$

with probability one. Therefore it suffices to prove

$$\int_{|t|>\eta n^{1/2}} |\chi_n(t)|dt \overset{a.s.}{\to} 0. \tag{A.2}$$

Notice that if N is standard normal, $|E\exp(isN+itN^2)|^5$ is integrable in (s,t). From this fact, taking $n \ge 5$, and letting ξ_n be the empiric characteristic function of the sample of pairs (X_i,X_i^2), $1 \le i \le n$, we see that the left

side of (A.2) is dominated by

$$\text{const. } n^{4\ell} \sup_{t_1^2+t_2^2>n^2} |\xi_n(t_1 - 2t_2\bar{X}, t_2)|^{n-5}$$

for all sufficiently large n. Finally observe that $\sup |\xi_n - E\xi_n| \overset{a.s.}{\to} 0$, and invoke condition (3.1).

REFERENCES

Babu, G.J. and Singh, K. (1983). Inference on means using the bootstrap. Annals of Statistics **11**, 999-1003.

Beran, R.J. (1982). Estimated sampling distributions: the bootstrap and competitors. Annals of Statistics **10**, 212-225.

Beran, R.J. (1984). Jackknife approximations to bootstrap estimates. Annals of Statistics **12**, 101-118.

Efron, B. (1982). The Jackknife, the Bootstrap and other Resampling Plans. SIAM, Philadelphia.

Geary, R.C. (1947). Testing for normality. Biometrika **34**, 209-242.

Hall, P. (1984). On the bootstrap and confidence intervals. University of North Carolina, Department of Statistics Mimeo Series No. 1574.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. Annals of Statistics **9**, 1187-1195.

END

FILMED

10-85

DTIC