The Annals of Probability
1988, Vol. 16, No. 4, 1665–1684

RATE OF CONVERGENCE IN BOOTSTRAP APPROXIMATIONS

By Peter Hall

Australian National University

Let $X_1, \dots, X_n$ be independent and identically distributed random variables with zero mean and unit variance. It is shown that the random bootstrap approximation to the distribution of $S \equiv n^{-1/2}\Sigma_j X_j$, converges to normality at precisely the same rate as $n^{-3/2}|\Sigma_j X_j^3| + n^{-2}\Sigma_j X_j^4$ converges to 0, up to terms of smaller order than $n^{-1/2}$. This result is used to explore properties of the bootstrap approximation under conditions weaker than existence of finite third moment. In most cases of that type it turns out that the bootstrap approximation to the distribution of S is asymptotically *equivalent* to the normal approximation, so that the numerical expense of calculating the bootstrap approximation would not be justified. There also exist circumstances where the third moment is "almost" finite, yet the bootstrap approximation is asymptotically much worse than the simpler normal approximation. Necessary and sufficient conditions are given for a one-term Edgeworth expansion of the bootstrap approximation.

1. Introduction. In sufficiently regular cases, the resampling or "bootstrap" approximation to an unknown distribution function has been established as an improvement over the simpler normal approximation [e.g., Efron (1981), Beran (1984) and Efron and Tibshirani (1986)]. In particular, it is known that if third moments are finite, then the bootstrap approximation to the distribution of a sum of independent random variables corrects for the skewness term of order $n^{-1/2}$ in an Edgeworth expansion [Bickel and Freedman (1980) and Singh (1981)], and thus betters the normal approximation. In the present paper we show that if third moments are infinite, very different behaviour is evident. Far from improving on the normal approximation, the bootstrap approximation can actually do worse than the normal approximation when third moments are infinite. In a great many cases with infinite third moment, the bootstrap approximation is asymptotically equivalent to the normal approximation, and there the numerical expense of computing the bootstrap approximation would not be justified.

We derive our results by studying the rate of convergence of the bootstrap approximation. Before giving more details, it is necessary to introduce notation. Let $X, X_1, X_2, \dots$ be independent and identically distributed random variables with finite variance, and assume for the sake of definiteness that $E(X^2) = 1$ and $E(X) = 0$. This standardization is not strictly necessary when studying the bootstrap, which is scale and location invariant, but it is nevertheless convenient. Write $\mathscr{X}$ for the collection $\{X_1, \dots, X_n\}$ and let $\{X_1^*, \dots, X_n^*\}$ be a collection

Received February 1987; revised February 1988.

AMS 1980 *subject classifications*. Primary 60F05, 60G50; secondary 62G99.

Key words and phrases. Bootstrap approximation, central limit theorem, Edgeworth expansion, normal approximation, rate of convergence.

drawn at random from $\mathscr{X}$, sampling with replacement. Put

$$\overline{X} \equiv n^{-1} \sum_{j=1}^{n} X_j, \qquad \overline{X}^* \equiv n^{-1} \sum_{j=1}^{n} X_j^*, \qquad \hat{\sigma}^2 \equiv n^{-1} \sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2,$$

$$S \equiv n^{1/2}\overline{X}, \qquad S^* \equiv n^{1/2}(\overline{X}^* - \overline{X})/\hat{\sigma},$$

$$F_n(x) \equiv P(S \le x), \quad \text{and} \quad F_n^*(x) \equiv P(S^* \le x|\mathscr{X}).$$

Let Φ and ϕ denote standard normal distribution and density functions, respectively.

The aim of much statistical theory is to approximate the unknown distribution function F_n. The normal approximation declares that when n is moderate to large, F_n may be approximated by Φ. The bootstrap approximation argues that F_n should be close to F_n^*. An important question to be resolved is: "When is F_n closer to F_n^* than to Φ?" We shall show that for all practical purposes the answer to this question is, "When X has finite *third* moment." Athreya (1985, 1987) has shown that, roughly speaking, finite *second* moments are required for *consistency* of the bootstrap approximation. See also Bickel and Freedman (1981).

Our statement that finiteness of the third moment is required for the bootstrap approximation to improve on the normal approximation, is an oversimplification. Nevertheless it is quite close to the truth. To delve a little deeper into the problem, let us consider the case

$$P(|X| > x) = x^{-\alpha}K(x), \tag{1.1}$$

where $2 \le \alpha \le 3$ and K is slowly varying at infinity. Assume that the tails of X are balanced, in the sense that $P(X > x)/P(|X| > x)$ converges as $x \to \infty$. When $2 \le \alpha < 3$, the normal approximation and bootstrap approximation are asymptotically equivalent. In fact for all but at most a finite number of values of x,

$$\lim_{n\to\infty} |F_n(x) - F_n^*(x)|/|F_n(x) - \Phi(x)| = 1 \tag{1.2}$$

almost surely. (See Examples 3.1 and 3.2 in Section 3.) The case $\alpha = 3$ is more tricky and depends very much on the choice of K. On some occasions [e.g., when X is symmetric and $K(x) \equiv \exp\{(\log x)^{\beta}\}$, some $0 < \beta < 1$], result (1.2) continues to hold. On others [e.g., X symmetric and $K(x) \equiv (\log x)^{\beta}$, some $\beta > 0$], we have instead

$$\limsup_{n\to\infty} |F_n(x) - F_n^*(x)|/|F_n(x) - \Phi(x)| = +\infty \tag{1.3}$$

almost surely, for all but at most a finite number of values of x. In this circumstance, the bootstrap approximation is asymptotically worse, along a subsequence, than the normal approximation. (See Example 3.3 for details.) In still other cases [e.g., when K is chosen so that $E(|X|^3) < \infty$ and when X is nonlattice and $E(X^3) \ne 0$],

$$\lim_{n\to\infty} |F_n(x) - F_n^*(x)|/|F_n(x) - \Phi(x)| = 0$$

almost surely for all $x \neq \pm 1$. Here the bootstrap is asymptotically better than the normal approximation. (See Example 3.5 in Section 3.) Thus, the case $\alpha = 3$ (corresponding to the third moment being either "just finite" or "just infinite") forms the boundary between circumstances where the bootstrap performs better than the simple normal approximation and circumstances where it does not.

In this simplified discussion we have assessed quality of the bootstrap approximation in "strong" terms, using almost sure convergence throughout. Our conclusions do sometimes change when quality is measured "weakly", using convergence in probability. Examples 3.4 and 3.5 in Section 3 will discuss this phenomenon.

Next we say a little about techniques used to derive our conclusions. We begin in Section 2 by studying the general problem of the rate of convergence of F_n^* to Φ. It turns out that with probability 1 and up to terms of smaller order than $n^{-1/2}$, the rate of convergence is *precisely* that of

$$\hat{\delta}_n \equiv n^{-3/2}\left|\sum_{j=1}^{n} X_j^3\right| + n^{-2}\sum_{j=1}^{n} X_j^4$$

to 0. [The only regularity condition needed for this result is $E(X^2) < \infty$. However, a continuity correction is required in the lattice case in order to remove the rounding-error term of order $n^{-1/2}$.] Now, $\hat{\delta}_n$ is a particularly simple function of sums of independent random variables and may be studied quite easily using standard results about such sums. That development is undertaken in Section 3. We obtain the limit theorem (1.2) when $\hat{\delta}_n$ is of smaller order than the difference between F_n and Φ; that is, when F_n^* converges to Φ more rapidly than does F_n. Proofs of main theorems are given in Section 4.

This approach to studying the bootstrap is different from those traditionally employed in problems of the same type, in that it does not require higher-order moment assumptions at the outset; compare Singh (1981) and Babu and Singh (1983, 1984). Our technique permits a detailed analysis of the bootstrap approximation in cases which have not been treated before. It also allows us to give extra detail in cases which have already been studied. In particular, Example 3.5 in Section 3 shows that if approximations are assessed using strong convergence, then a *necessary and sufficient* condition for the bootstrap to correct for the skewness term of order $n^{-1/2}$ in an Edgeworth expansion is that the third moment be finite. Singh (1981) showed that finite third moment is *sufficient*. We also produce a condition weaker than the existence of finite third moment, which is necessary and sufficient for Edgeworth correction if approximations are assessed using convergence in probability rather than almost sure convergence.

2. Leading term for random approximation. The leading term in an expansion of the "normal error" $P(S \leq x) - \Phi(x)$ is

$$\begin{aligned} L_n(x) &\equiv nE\{\Phi(x - n^{-1/2}X)\} - n\Phi(x) - \tfrac{1}{2}\phi'(x) \\ &= nE\Big\{\Phi(x - n^{-1/2}X) - \Phi(x) + n^{-1/2}X\phi(x) \\ &\qquad - \tfrac{1}{2}\big(n^{-1/2}X\big)^2\phi'(x)\Big\}. \end{aligned} \tag{2.1}$$

See Hall [(1982), Chapter 2, Section 2.3]. By analogy, the leading term in an expansion of $P(S^* \le x|\mathscr{X}) - \Phi(x)$ is

$$\tilde{L}_n(x) \equiv \sum_{j=1}^{n} \Phi\left\{x - n^{-1/2}\hat{\sigma}^{-1}\left(X_j - \overline{X}\right)\right\} - n\Phi(x) - \tfrac{1}{2}\phi'(x).$$

We shall show in Section 4 during our proof of Theorem 2.2 that $\tilde{L}_n$ is asymptotically equivalent to

$$\begin{aligned} \hat{L}_n(x) &\equiv \sum_{j=1}^{n} \Phi\left(x - n^{-1/2}X_j\right) - n\Phi(x) + n^{1/2}\overline{X}\phi(x) - \tfrac{1}{2}\left(n^{-1}\sum_{j=1}^{n} X_j^2\right)\phi'(x) \\ &= \sum_{j=1}^{n} \left\{\Phi\left(x - n^{-1/2}X_j\right) - \Phi(x) + n^{-1/2}X_j\phi(x) - \tfrac{1}{2}\left(n^{-1/2}X_j\right)^2\phi'(x)\right\}. \end{aligned} \tag{2.2}$$

Since $\hat{L}_n$ does not involve $\hat{\sigma}$ and so is superficially a little simpler than $\tilde{L}_n$, we shall state our results for $\hat{L}_n$ rather than $\tilde{L}_n$.

The order of magnitude of L_n is that of

$$\begin{aligned} \delta_n \equiv E\left\{X^2 I(|X| > n^{1/2})\right\} &+ n^{-1/2}\left|E\left\{X^3 I(|X| \le n^{1/2})\right\}\right| \\ &+ n^{-1}E\left\{X^4 I(|X| \le n^{1/2})\right\}, \end{aligned} \tag{2.3}$$

irrespective of whether distance is measured in integral or supremum metrics. For example, defining $l_n \equiv \sup|L_n|$ we have

$$C_1\delta_n \le l_n \le C_2\delta_n \tag{2.4}$$

for constants $C_2 > C_1 > 0$, the constant C_2 being absolute. See Hall [(1982), Chapter 2, Section 2.4]. The order of magnitude of $\hat{L}_n$ and of $\tilde{L}_n$ is that of

$$\hat{\delta}_n \equiv n^{-3/2}\left|\sum_{j=1}^{n} X_j^3\right| + n^{-2}\sum_{j=1}^{n} X_j^4, \tag{2.5}$$

in a wide variety of different metrics. For example, we have

THEOREM 2.1. (i) *There exists an absolute constant* $C_2 > 0$ *such that for all samples* $\mathscr{X}$,

$$\sup_{-\infty < x < \infty} |\hat{L}_n(x)| \le C_2\hat{\delta}_n.$$

(ii) *Let* x_1 *be any element of the set* $\{-1, 1\}$ *and* x_2 *any real number not in that set. If* $E(X^2) < \infty$ *then there exists a constant* $C_1 = C_1(x_2) > 0$ *such that*

$$P\left\{\sup_{x \in \{x_1, x_2\}} |\hat{L}_n(x)| \ge C_1\hat{\delta}_n \text{ for all } n \ge m\right\} \to 1$$

as $m \to \infty$.

Define $\hat{l}_n \equiv \sup|\hat{L}_n|$. An immediate corollary of Theorem 2.1 is that with probability 1,

$$C_1 \leq \liminf_{n\to\infty}\left(\hat{l}_n/\hat{\delta}_n\right) \leq \limsup_{n\to\infty}\left(\hat{l}_n/\hat{\delta}_n\right) \leq C_2 \tag{2.6}$$

[compare (2.4)], where $C_1 \equiv \sup_x C_1(x)$.

It is worth making the trite remark that if $E(X^2) < \infty$, then $\delta_n \to 0$ and $\hat{\delta}_n \to 0$ almost surely. In the case of δ_n, this fact is trivial and well known. For $\hat{\delta}_n$ it follows indirectly from Bickel and Freedman (1981), but may be proved directly as follows. Observe that for any $0 < \eta < 1$ and with probability 1, the event $\mathscr{E}_n \equiv \{|X_j| \leq \eta n^{1/2}, 1 \leq j \leq n\}$ holds for all sufficiently large n. On $\mathscr{E}_n$ we have $\hat{\delta}_n \leq 2\eta n^{-1}\Sigma_j X_j^2 \to 2\eta$ almost surely, and η may be chosen arbitrarily small.

We are now in a position to describe the size of $P(S^* \leq x|\mathscr{X}) - \Phi(x) - \hat{L}_n(x)$. We wish to go as far as terms which are of smaller order than both $\hat{\delta}_n$ and $n^{-1/2}$. However, if the underlying distribution of X is lattice, then rounding-error terms of order $n^{-1/2}$ appear in expansions of $P(S^* \leq x|\mathscr{X})$, and we should remove them first. Suppose X is lattice and takes only values of the form $a + jb$ ($j = 0, \pm 1, \pm 2, \dots$), where $b > 0$ is the maximal span of the lattice. Let $R(x) \equiv \langle x\rangle - x + \frac{1}{2}$, where $\langle x\rangle$ denotes the integer part of x, and put

$$\hat{R}_n(x) \equiv R(n^{1/2}\hat{\sigma}x/b)\phi(x)b/(n^{1/2}\hat{\sigma}).$$

Define $\hat{R}_n \equiv 0$ if the distribution of X is nonlattice. Let

$$\hat{\Delta}_n \equiv \sup_{-\infty<x<\infty} |P(S^* \leq x|\mathscr{X}) - \Phi(x) - \hat{L}_n(x) - \hat{R}_n(x)|.$$

THEOREM 2.2. *If $E(X^2) < \infty$, then $\hat{\Delta}_n/(\hat{\delta}_n + n^{-1/2}) \to 0$ almost surely as $n \to \infty$.*

In the case where X has a lattice distribution, we actually prove a good deal more than Theorem 2.2. We show that if $\hat{L}_n$ is replaced by $\tilde{L}_n$ in the definition of $\hat{\Delta}_n$, then $\hat{\Delta}_n/(\hat{\delta}_n^2 + n^{-1})$ is almost surely bounded as $n \to \infty$. See formula (4.8) in Section 4. This result may also be obtained in the nonlattice case, provided we assume that the characteristic function α of X satisfies Cramér's condition

$$\limsup_{|t|\to\infty} |\alpha(t)| < 1.$$

It is necessary only to modify an argument toward the end of our proof of Theorem 2.2.

It is of interest to compare Theorem 2.2 with its analogue for S. There we define the rounding-error correction to be

$$R_n(x) \equiv R\{(n^{1/2}x - na)/b\}\phi(x)b/n^{1/2}$$

in the lattice case, assuming $E(X^2) = 1$ and $E(X) = 0$; and $R_n \equiv 0$ in the nonlattice case. Recall that L_n is given by (2.1) and put

$$\Delta_n \equiv \sup_{-\infty<x<\infty} |P(S \leq x) - \Phi(x) - L_n(x) - R_n(x)|.$$

The following result is taken from Hall [(1982), Theorems 4.2 and 4.3, pages 162 and 164].

THEOREM 2.3. *If* $E(X^2) = 1$ *and* $E(X) = 0$, *then* $\Delta_n/(\delta_n + n^{-1/2}) \to 0$ *as* $n \to \infty$.

Comparing Theorems 2.2 and 2.3 we see that, provided we correct for rounding errors in the lattice case, the distance between the distribution functions $P(S^* \le \cdot|\mathscr{X})$ and Φ is of precise order $\hat{\delta}_n + o(n^{-1/2})$, whereas the distance between $P(S \le \cdot)$ and Φ is of precise order $\delta_n + o(n^{-1/2})$.

3. Accuracy of the bootstrap approximation. Let us assume for the sake of simplicity that $E(X^2) = 1$, $E(X) = 0$ and X has a nonlattice distribution. Then the error in the normal approximation to the distribution of S is

$$P(S \le x) - \Phi(x) = L_n(x) + r_n(x), \tag{3.1}$$

where $\sup|r_n| = o(\delta_n + n^{-1/2})$. See Theorem 2.3. The error in the bootstrap approximation to the distribution of S is

$$\begin{aligned} P(S \le x) - P(S^* \le x|\mathscr{X}) &= P(S \le x) - \Phi(x) - \{P(S^* \le x|\mathscr{X}) - \Phi(x)\} \\ &= L_n(x) - \hat{L}_n(x) + \hat{r}_{n1}(x), \end{aligned}$$

where $\sup|\hat{r}_{n1}| = o(\delta_n + \hat{\delta}_n + n^{-1/2})$ almost surely. See Theorems 2.2 and 2.3, and remember that $L_n, \hat{L}_n, \delta_n, \hat{\delta}_n$ were defined in (2.1)–(2.3) and (2.5). We showed in (2.4) and (2.6) that L_n is of precise order δ_n and $\hat{L}_n$ is of precise order $\hat{\delta}_n$, as measured by the supremum metric. Therefore, if $\hat{\delta}_n/\delta_n \to 0$ almost surely, then

$$P(S \le x) - P(S^* \le x|\mathscr{X}) = L_n(x) + \hat{r}_{n2}(x), \tag{3.2}$$

where $\sup|\hat{r}_{n2}| = o(\delta_n + n^{-1/2})$ almost surely. In this circumstance the bootstrap approximation is asymptotically equivalent to the normal approximation up to terms of order $n^{-1/2}$, as may be seen by comparing (3.1) and (3.2).

In the present section we argue that in a great many circumstances where $|X|$ has infinite third moment, the bootstrap approximation is asymptotically equivalent to the normal approximation in the sense just described. Indeed, this result is true if $P(|X| > x) = x^{-\alpha}K(x)$, where $2 \le \alpha < 3$ and the function K is slowly varying at infinity. [Under this condition, $E(|X|^{\alpha+\varepsilon}) = \infty$ and $E(|X|^{\alpha-\varepsilon}) < \infty$ for $0 < \varepsilon < \alpha$.] However, in the case $\alpha = 3$ it is possible for the bootstrap approximation to be asymptotically *worse* than the normal approximation, in the sense that $\limsup(\hat{\delta}_n/\delta_n) = +\infty$ and

$$\limsup_{n\to\infty} \frac{\sup_{-\infty<x<\infty}|P(S \le x) - P(S^* \le x|\mathscr{X})|}{\sup_{-\infty<x<\infty}|P(S \le x) - \Phi(x)|} = +\infty \tag{3.3}$$

almost surely.

These results will be obtained in a sequence of five examples, of which the first four treat the circumstance where $P(|X| > x) = x^{-\alpha}K(x)$ and K is slowly varying. Examples 3.1 and 3.2 deal with cases $\alpha = 2$ and $2 < \alpha < 3$, respectively; Examples 3.3 and 3.4 treat $\alpha = 3$; and Example 3.5 gives necessary and sufficient conditions for one-term Edgeworth corrections discussed by Singh (1981).

The following lemma on strong convergence is needed for the examples and will be proved in Section 4. Let $Y, Y_1, Y_2, \ldots$ be independent and identically distributed random variables and put $T_n \equiv \Sigma_{1 \le j \le n} Y_j$. A sequence of positive constants $\{b_n\}$ will be said to be *approximately increasing* if $\{nb_n\}$ is nondecreasing and if for some $C > 0$,

$$b_n \le C \inf_{k \ge n} b_k \quad \text{for all } n \ge 1. \tag{3.4}$$

LEMMA 3.1. *Assume either $E(|Y|) = \infty$ or $E(|Y|) < \infty$ and $E(Y) = 0$. If $\{n^{-1}c_n\}$ is approximately increasing, then*

$$\limsup_{n \to \infty} |T_n|/c_n \begin{cases} = 0 \\ > 0 \end{cases} \textit{almost surely according as} \quad \sum_{n=1}^{\infty} P(|Y| > c_n) \begin{cases} < \infty \\ = \infty \end{cases}.$$

The symbols $C, C_1, C_2, \ldots$ will denote positive generic constants. In Examples 3.1–3.4, we shall assume for simplicity that X has a nonlattice distribution.

EXAMPLE 3.1. Suppose $P(|X| > x) = x^{-2}K(x)$, where K is slowly varying at infinity. We shall prove that $\hat{\delta}_n/\delta_n \to 0$ almost surely, implying that the bootstrap approximation is asymptotically equivalent to the simpler normal approximation up to terms of order $n^{-1/2}$.

Integration by parts and application of Theorems 2.6 and 2.7 of Seneta (1976) show that each of $E\{X^2 I(|X| > n^{1/2})\}$, $n^{-1/2}E\{|X|^3 I(|X| \le n^{1/2})\}$ and $n^{-1}E\{X^4 I(|X| \le n^{1/2})\}$ is asymptotic to a (different) constant multiple of

$$\int_{n^{1/2}}^{\infty} x^{-1}K(x)\,dx.$$

Therefore, with

$$\delta_n' \equiv E\{X^2 \min(1, n^{-1}X^2)\} \sim C_1 \int_{n^{1/2}}^{\infty} x^{-1}K(x)\,dx,$$

we have $\delta_n' \le \delta_n \le C_2\delta_n'$. The sequence $\{n^{1/2}\delta_n'\}$ is approximately increasing; see Seneta [(1976), page 20]. Furthermore,

$$\begin{aligned}
\Sigma &\equiv \sum_{n=1}^{\infty} P\big(|X|^3 > n^{3/2}\delta_n'\big) \\
&\le C_3 \sum_{n=1}^{\infty} n^{-1}\left\{\int_{n^{1/2}}^{\infty} x^{-1}K(x)\,dx\right\}^{-2/3} K\left[n^{1/2}\left\{\int_{n^{1/2}}^{\infty} x^{-1}K(x)\,dx\right\}^{1/3}\right] \\
&\le C_4 \int_1^{\infty} y^{-1}\left\{\int_y^{\infty} x^{-1}K(x)\,dx\right\}^{-2/3} K\left[y\left\{\int_y^{\infty} x^{-1}K(x)\,dx\right\}^{1/3}\right] dy.
\end{aligned}$$

It follows from Lemma 3.2 that for any $\xi > 0$,

$$K\left[y\left\{\int_y^\infty x^{-1}K(x)\,dx\right\}^{1/3}\right] \le C_5(\xi)K(y)\left\{\int_y^\infty x^{-1}K(x)\,dx\right\}^{-\xi}$$

for $y \ge 1$. Taking $\xi < \frac{1}{3}$ we conclude that

$$\sum \le C_4C_5(\xi)\int_1^\infty y^{-1}K(y)\left\{\int_y^\infty x^{-1}K(x)\,dx\right\}^{-(2/3)-\xi} dy < \infty.$$

From this result and Lemma 3.1 we see that

$$\left(n^{3/2}\delta_n\right)^{-1}\sum_{j=1}^n |X_j|^3 \to 0$$

almost surely. It then follows that $(n^{-2}\Sigma_j X_j^4)/\delta_n \to 0$ almost surely, for under the condition $E(X^2) < \infty$ the event $\mathscr{E}_n \equiv \{|X_j| \le n^{1/2},\ 1 \le j \le n\}$ holds for all sufficiently large n, with probability 1, and on $\mathscr{E}_n$ we have $n^{-2}\Sigma_j X_j^4 \le n^{-3/2}\Sigma_j|X_j|^3$. Therefore, $\hat{\delta}_n/\delta_n \to 0$ almost surely, as had to be proved.

The following lemma is easily proved via Karamata's representation theorem [Seneta (1976), Theorem 1.2, page 2].

LEMMA 3.2. *If K is slowly varying at infinity and if $\varepsilon = \varepsilon(x) \to 0$ as $x \to \infty$ in such a manner that $x\varepsilon(x) \to \infty$, then for each $\xi > 0$, $\varepsilon^\xi K(\varepsilon x)/K(x) \to 0$.*

It may be shown that in the case of Example 3.1,

$$L_n(x) = -\tfrac{1}{2}\phi'(x)E\left\{X^2 I(|X| > n^{1/2})\right\} + s_n(x),$$

where

$$\sup|s_n| = o\left[E\left\{X^2 I(|X| > n^{1/2})\right\}\right] = o(\delta_n);$$

see Höglund (1970) and Hall [(1982), Theorem 4.10(ii), page 196]. Therefore, result (1.2) holds for all x except possibly $x = 0$.

EXAMPLE 3.2. Suppose $C_1x^{-\beta} \le P(|X| > x) \le C_2x^{-\gamma}$ for all $x \ge 1$, where $0 < \gamma \le \beta < \infty$ and $(5-\beta)\gamma > 6$. These specifications include the case where $P(|X| > x) = x^{-\alpha}K(x)$, $2 < \alpha < 3$ and K is slowly varying, for there we may select $\beta \in (\alpha, 3)$ and $\gamma \in (2, \alpha)$ such that $(5-\beta)\gamma > 6$. We shall prove that $\hat{\delta}_n/\delta_n \to 0$ almost surely, implying that the bootstrap approximation is asymptotically equivalent to the normal approximation up to terms of order $n^{-1/2}$.

Notice that $\delta_n \ge nP(|X| > n^{1/2}) \ge C_1n^{1-(\beta/2)}$ and

$$\sum_{n=1}^\infty P\left(|X|^3 > n^{3/2}n^{1-(\beta/2)}\right) = \sum_{n=1}^\infty P\left(|X| > n^{(5-\beta)/6}\right) \le C_2\sum_{n=1}^\infty n^{-(5-\beta)\gamma/6} < \infty.$$

Therefore, by Lemma 3.1, $(n^{-3/2}\Sigma_j|X_j|^3)/\delta_n \to 0$ almost surely, which implies (as in Example 3.1) that $(n^{-2}\Sigma_j X_j^4)/\delta_n \to 0$ almost surely. Hence $\hat{\delta}_n/\delta_n \to 0$ almost surely, as had to be shown.

If $P(|X| > x) = x^{-\alpha}K(x)$ where $2 < \alpha < 3$ and K is slowly varying at infinity, and if the tails of X are balanced in the sense that $P(X > x)/P(|X| > x)$ converges as $x \to \infty$, then

$$L_n(x) = A(x)nP(|X| > n^{1/2}) + s_n(x),$$

where $\sup|s_n| = o(\delta_n)$. The function A does not depend on n, has at most four zeros and is given by Höglund (1970) and Hall [(1982), Theorem 4.10(i), page 196]. Result (1.2) holds if x is not a zero of A.

EXAMPLE 3.3. Suppose $P(|X| > x) = x^{-3}K(x)$, where K is slowly varying at infinity, that $E(|X|^3) = \infty$ and that X is symmetric. [The case $E(|X|^3) < \infty$ will be treated in Example 3.5.] We shall show that, depending on choice of K, it is possible to have either $\hat{\delta}_n/\delta_n \to 0$ almost surely or $\limsup \hat{\delta}_n/\delta_n = +\infty$ almost surely. Therefore, the bootstrap approximation can be equivalent to the normal approximation up to terms of order $n^{-1/2}$, or worse than the normal approximation along a subsequence.

Notice that $C_1 n^{-1/2}K(n^{1/2}) \le \delta_n = E\{X^2 \min(1, n^{-1}X^2)\} \le C_2 n^{-1/2}K(n^{1/2})$. We first prove that

$$\left(n^{-2} \sum_{j=1}^{n} X_j^4\right) \Big/ \delta_n \to 0 \tag{3.5}$$

almost surely. For any $\varepsilon > 0$,

$$\begin{aligned}\sum_{n=1}^{\infty} P\left(X^4 > n^2\delta_n\right) &\le C_3 \sum_{n=1}^{\infty} \{n^{3/2}K(n^{1/2})\}^{-3/4} K\left[\{n^{3/2}K(n^{1/2})\}^{1/4}\right] \\ &\le C_4(\varepsilon) \sum_{n=1}^{\infty} n^{-(9/8)+\varepsilon},\end{aligned}$$

which converges if ε is sufficiently small. The sequence $\{n\delta_n\}$ is approximately increasing and so (3.5) follows from Lemma 3.1.

In view of (3.5), the efficacy of the bootstrap approximation depends on behaviour of $(n^{-3/2}|\sum_j X_j^3|)/\delta_n$ as $n \to \infty$. If $K(x) \equiv (\log x)^\alpha$ for $\alpha \ge -1$, then it follows from result (4.18) of Kesten (1972) that $\limsup(n^{-3/2}|\sum_j X_j^3|)/\delta_n = +\infty$ and so $\limsup \hat{\delta}_n/\delta_n = +\infty$. When $\alpha > 0$ this means that $\limsup \hat{\delta}_n/(\delta_n + n^{-1/2}) = +\infty$, so that the bootstrap approximation is worse (along a subsequence) than the normal approximation. On the other hand, if $K(x) \equiv \exp\{(\log x)^\alpha\}$ for some $\alpha \in (0,1)$, then $\hat{\delta}_n/\delta_n \to 0$ almost surely. To see why, notice that for any $\varepsilon > 0$,

$$\begin{aligned}\sum_{n=1}^{\infty} P\left(|X|^3 > n^{3/2}\delta_n\right) &\le C_5 \int_1^{\infty} \{xK(x)\}^{-1} K\left\{x^{2/3}K(x)^{1/3}\right\} dx \\ &\le C_6(\varepsilon) \int_1^{\infty} x^{-1} \exp\left[-\left\{1 - (2/3)^\alpha - \varepsilon\right\}(\log x)^\alpha\right] dx,\end{aligned}$$

which converges if ε is sufficiently small. It now follows via Lemma 3.1 that $(n^{-3/2}|\sum X_j^3|)/\delta_n \to 0$ and so $\hat{\delta}_n/\delta_n \to 0$, almost surely. Therefore, the bootstrap approximation is equivalent to the normal approximation up to terms of order $n^{-1/2}$.

EXAMPLE 3.4. In Examples 3.1–3.3 we assessed quality of the bootstrap approximation in "strong" terms, basing our arguments on almost sure convergence. We could have looked instead at convergence in probability. In the majority of cases (e.g., Examples 3.1 and 3.2) this would have made no difference, since the most common occurrence was $\hat{\delta}_n/\delta_n \to 0$ almost surely, in which case the bootstrap approximation was asymptotically equivalent to the normal approximation (up to terms of order $n^{-1/2}$) both weakly and strongly. However there do exist circumstances where $\limsup \hat{\delta}_n/\delta_n = +\infty$ almost surely (meaning that in a strong sense, the bootstrap is much worse than the normal approximation along a subsequence), but $\hat{\delta}_n/\delta_n = O(1)$ in probability (so that in a weak sense, bootstrap and normal approximations are roughly equivalent). Thus we can draw quite different conclusions if we assess the bootstrap approximation in weak and strong terms.

A case in point is given by Example 3.3 with $K(x) \equiv (\log x)^\alpha$, any $\alpha > 0$; there we have $\limsup \hat{\delta}_n/(\delta_n + n^{-1/2}) = +\infty$ almost surely, but $\hat{\delta}_n/\delta_n = O(1)$ in probability. In this circumstance, result (1.3) follows via Theorem 1 of Höglund (1970). We shall give a more curious example, where the random approximation supplied by the bootstrap is asymptotically equivalent (in a weak sense) to a nonrandom approximation but not to the normal approximation, and where $\limsup \hat{\delta}_n/(\delta_n + n^{-1/2}) = +\infty$ almost surely yet $\hat{\delta}_n/\delta_n = O_p(1)$.

Take $X \equiv Y - E(Y)$, where Y is nonnegative and $P(Y > y) = y^{-3}$ for large y. Then $\delta_n \sim Cn^{-1/2}\log n$ and result (3.5) follows as before. The series $\sum_n P(|X|^3 > \lambda n \log n)$ diverges for each $\lambda > 0$ and so by Feller [(1946), Theorem 2] or Stout [(1974), Theorem 3.2.5, page 132], $\limsup(n^{-3/2}|\sum_j X_j^3|)/\delta_n = +\infty$ almost surely. Therefore, $\limsup \hat{\delta}_n/\delta_n = +\infty$. By Klass and Teicher [(1977), Theorem 1],

$$\left(n^{3/2}\delta_n\right)^{-1} \sum_{j=1}^{n} \left[X_j^3 - E\{X^3 I(|X| \le n^{1/3})\}\right] \to 0$$

in probability. Now,

$$\left|\hat{L}_n(x) - \tfrac{1}{6}(-n^{-1/2})^3\left(\sum_{j=1}^{n} X_j^3\right)\phi''(x)\right| \le \tfrac{1}{24}(\sup|\phi'''|)n^{-2}\sum_{j=1}^{n} X_j^4,$$

and the right-hand side equals $o(\delta_n)$ almost surely. Therefore, with

$$L_n^\dagger(x) \equiv -\tfrac{1}{6}n^{-1/2}E\{X^3 I(|X| \le n^{1/3})\}\phi''(x),$$

we have

$$\sup_{-\infty < x < \infty} |\hat{L}_n(x) - L_n^\dagger(x)| = o(\delta_n)$$

in probability. Consequently, in a weak sense the bootstrap approximation is asymptotically equivalent to nonrandom approximation by $\Phi + L_n^\dagger$:

$$P(S^* \le x|\mathscr{X}) = \Phi(x) + L_n^\dagger(x) + o_p(n^{-1/2}\log n).$$

The normal approximation declares that

$$P(S \le x) = \Phi(x) + L_n(x) + o(n^{-1/2}\log n).$$

Both L_n and $L_n^\dagger$ are $O(n^{-1/2}\log n)$, but these functions are not particularly close. Indeed for any x,

$$\begin{aligned} &L_n^\dagger(x) - L_n(x) \\ &\quad = \tfrac{1}{6}n^{-1/2}E\{X^3I(n^{1/3} < |X| \le n^{1/2})\}\phi''(x) + O(n^{-1/2}) \end{aligned} \tag{3.6}$$

and $E\{X^3I(n^{1/3} < |X| \le n^{1/2})\} \sim \frac{1}{2}\log n$ as $n \to \infty$. [Result (3.6) follows from the facts

$$\begin{aligned} &\left|L_n(x) + \tfrac{1}{6}n^{-1/2}E\{X^3I(|X| \le n^{1/2})\}\phi''(x)\right| \\ &\quad \le C_1\left[E\{X^2I(|X| > n^{1/2})\} + n^{-1}E\{X^4I(|X| \le n^{1/2})\}\right] \\ &\quad = C_1E\{X^2\min(1, n^{-1}X^2)\}, \end{aligned}$$

true for all x, and

$$E\{X^2\min(1, n^{-1}X^2)\} \le C_2\int_1^\infty x^{-2}\min(1, n^{-1}x^2)\,dx = O(n^{-1/2}).]$$

EXAMPLE 3.5. Here we use Theorem 2.2 to deduce necessary and sufficient conditions for a one-term Edgeworth expansion. Let β be any given constant; it might be $E(X^3)$ if that moment were finite. Define $p(x) \equiv -(\beta/6)(x^2 - 1)$ and

$$\hat{\gamma}_n \equiv \sup_{-\infty < x < \infty} \left|P(S^* \le x|\mathscr{X}) - \{\Phi(x) + n^{-1/2}p(x)\phi(x) + \hat{R}_n(x)\}\right|.$$

[We include $\hat{R}_n(x)$ just in case X has a lattice distribution. In this example we do not exclude the lattice case.] We shall prove

THEOREM 3.1. (i) $\hat{\gamma}_n = o(n^{-1/2})$ *almost surely if and only if* $E(|X|^3) < \infty$ *and* $E(X^3) = \beta$.

(ii) $\hat{\gamma}_n = o(n^{-1/2})$ *in probability if and only if, as* $x \to \infty$, $xP(|X|^3 > x) \to 0$ *and* $E\{X^3I(|X| \le x)\} \to \beta$.

The conditions on X in (ii) are of course satisfied if $E(|X|^3) < \infty$ and $E(X^3) = \beta$. They also hold in other cases; for example, they hold with $\beta = 0$ if X is symmetric with $P(|X|^3 > x) \sim x^{-1}(\log x)^{-\alpha}$, for $\alpha > 0$. Only in the case $\alpha > 1$ does the latter example satisfy $E(|X|^3) < \infty$.

To interpret statements (i) and (ii) it is helpful to define

$$\gamma_n \equiv \sup_{-\infty < x < \infty} \left| P(S \le x) - \left\{ \Phi(x) + n^{-1/2} p(x)\phi(x) + R_n(x) \right\} \right|$$

and notice that necessary and sufficient conditions for $\gamma_n = o(n^{-1/2})$ are $xE\{X^2 I(|X| > x)\} \to 0$ and $E\{X^3 I(|X| \le x)\} \to \beta$. [See, for example, Hall (1982), page 186.] These conditions are strictly stronger than those given in statement (ii), but strictly weaker than $E(|X|^3) < \infty$ and $E(X^3) = \beta$. They are satisfied (with $\beta = 0$) in the case of the example in the previous paragraph.

PROOF OF THEOREM 3.1. Define

$$\hat{\varepsilon}_n \equiv n^{-3/2} \left| \sum_{j=1}^{n} \left(X_j^3 - \beta \right) \right| + n^{-1} \sum_{j=1}^{n} X_j^4$$

and

$$\hat{k}_n \equiv \sup_{-\infty < x < \infty} |\hat{L}_n(x) - p(x)\phi(x)|.$$

The argument used to derive Theorem 2.1 may be trivially modified to prove the existence of constants $C_2 > C_1 > 0$ such that with probability 1,

$$C_1 \le \liminf_{n \to \infty} \left(\hat{k}_n / \hat{\varepsilon}_n \right) \le \limsup_{n \to \infty} \left(\hat{k}_n / \hat{\varepsilon}_n \right) \le C_2;$$

compare (2.6). Therefore, by Theorem 2.2, $\hat{\gamma}_n = o(n^{-1/2})$ almost surely if and only if $n^{1/2}\hat{\varepsilon}_n \to 0$ almost surely. The latter condition implies

$$n^{-1} \sum_{j=1}^{n} \left(X_j^3 - \beta \right) \to 0 \tag{3.7}$$

almost surely, for which it is necessary and sufficient that $E(|X|^3) < \infty$ and $E(X^3) = \beta$ [see Chung (1974), page 126]. And if $E(|X|^3) < \infty$, then $n^{-3/2} \sum_j X_j^4 \to 0$ almost surely, since for any $\eta > 0$ and all sufficiently large n,

$$n^{-3/2} \sum_{j=1}^{n} X_j^4 = n^{-3/2} \sum_{j=1}^{n} X_j^4 I\left(|X_j| \le \eta n^{1/2}\right) \le n^{-1}\eta \sum_{j=1}^{n} |X_j|^3 \to \eta E\left(|X|^3\right).$$

This proves statement (i).

To derive (ii), note that by Theorem 2.2, $\hat{\gamma}_n = o(n^{-1/2})$ in probability if and only if $n^{1/2}\hat{\varepsilon}_n \to 0$ in probability. This condition implies that (3.7) holds with convergence in probability, for which the conditions in statement (ii) are necessary and sufficient [see Gnedenko and Kolmogorov (1968), page 134]. From those conditions it follows that $E(|X|^{3-\varepsilon}) < \infty$ whenever $0 < \varepsilon < 3$, and thence that $n^{-3/2} \sum_j X_j^4 \to 0$ in probability, since for any $\eta > 0$,

$$\begin{aligned} P\left(n^{-3/2} \sum_{j=1}^{n} X_j^4 > \eta \right) &\le nP\left(|X| > n^{1/3}\right) + P\left(n^{-3/2} n^{5/12} \sum_{j=1}^{n} |X|^{4-(5/4)} > \eta \right) \\ &\le nP\left(|X|^3 > n\right) + \left(\eta n^{13/12}\right)^{-1} nE\left(|X|^{11/4}\right) = o(1). \qquad \square \end{aligned}$$

4. Proofs. The symbols $C, C_1, C_2, \ldots$ will denote positive generic constants. It is convenient to prove Theorem 2.2 before Theorem 2.1. For convenience we assume throughout that $E(X) = 0$ and $E(X^2) = 1$.

PROOF OF THEOREM 2.2. Let $\hat{\alpha}_n$ denote the empirical characteristic function of the "sample" $\{(X_j - \bar{X})/\hat{\sigma}, 1 \le j \le n\}$. That is,

$$\hat{\alpha}_n(t) = n^{-1} \sum_{j=1}^{n} \exp\{it(X_j - \bar{X})/\hat{\sigma}\}.$$

We begin with a lemma.

LEMMA 4.1. *Assume* $E(X) = 0$ *and* $E(X^2) = 1$. *Given* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that with probability* 1,

$$\limsup_{n \to \infty} \sup_{0 < |t| \le \delta} t^{-2} |\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2| \le \varepsilon.$$

PROOF. Since $|e^{ix} - 1 - ix + \frac{1}{2}x^2| \le C \min(x^2, |x|^3)$ uniformly in x, then for $0 < \eta \le 1$ and $0 < t \le \delta$,

$$\begin{aligned}
&t^{-2}|\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2| \\
&\quad = t^{-2}n^{-1}\left| \sum_{j=1}^{n} \left[\exp\{it\hat{\sigma}^{-1}(X_j - \bar{X})\} - 1 - it\hat{\sigma}^{-1}(X_j - \bar{X}) \right.\right. \\
&\qquad\qquad \left.\left. + \tfrac{1}{2}t^2\hat{\sigma}^{-2}(X_j - \bar{X})^2\right]\right| \\
&\quad \le Ctn^{-1} \sum_{j=1}^{n} \hat{\sigma}^{-3}|X_j - \bar{X}|^3 I(t\hat{\sigma}^{-1}|X_j - \bar{X}| \le \eta) \\
&\qquad + Cn^{-1} \sum_{j=1}^{n} \hat{\sigma}^{-2}(X_j - \bar{X})^2 I(t\hat{\sigma}^{-1}|X_j - \bar{X}| > \eta) \\
&\quad \le C\left\{\eta + n^{-1} \sum_{j=1}^{n} \hat{\sigma}^{-2}(X_j - \bar{X})^2 I(\hat{\sigma}^{-1}|X_j - \bar{X}| > \eta/\delta)\right\}.
\end{aligned}$$

The right-hand side does not depend on t and converges almost surely to

$$C[\eta + E\{X^2 I(|X| > \eta/\delta)\}]$$

as $n \to \infty$. The latter may be made arbitrarily small by choosing η small and then δ small. This proves the lemma if the inner supremum is taken over $0 < t \le \delta$; the case $-\delta \le t < 0$ is similar. □

Observe that

$$\hat{\alpha}_n(t)^n = \exp[n \log\{1 + \hat{\alpha}_n(t) - 1\}] = \exp[n\{\hat{\alpha}_n(t) - 1\} + n\hat{r}_{n1}(t)],$$

where, arguing as in Hall [(1982), page 14], $|\hat{r}_{n1}(t)| \le \frac{1}{2}|\hat{\alpha}_n(t) - 1|^2/\{1 - |\hat{\alpha}_n(t) - 1|\}$. Also, since $|e^{is} - 1 - is| \le \frac{1}{2}s^2$ for all real s, then $|\hat{\alpha}_n(t) - 1| \le \frac{1}{2}t^2$ uniformly in t [so that $|\hat{\alpha}_n(t) - 1| \le \frac{1}{2}$ for $|t| \le 1$] and

$$\begin{aligned}\hat{\alpha}_n(t)^n &= \exp\left[n\left\{\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2\right\} + n\hat{r}_{n1}(t)\right]e^{-nt^2/2} \\ &= \left[1 + n\left\{\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2\right\}\right]e^{-nt^2/2} + \hat{r}_{n2}(t),\end{aligned} \tag{4.1}$$

where, if $|t| \le 1$,

$$\begin{aligned}|\hat{r}_{n2}(t)| \le C_1&\left\{n^2|\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2|^2 + n|\hat{\alpha}_n(t) - 1|^2 + n^2|\hat{\alpha}_n(t) - 1|^4\right\} \\ &\times\exp\left\{n|\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2| + n|\hat{\alpha}_n(t) - 1|^2 - nt^2/2\right\}.\end{aligned}$$

[Here we have used the fact that

$$|e^{z+w} - 1 - z| \le C_1(|z|^2 + |w|^2 + |w|)\exp(|z| + |w|).]$$

In view of Lemma 4.1, we may choose $\delta \in (0, \frac{1}{2})$ so small that

$$\limsup_{n\to\infty} \sup_{0<|t|\le\delta} t^{-2}|\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2| \le 1/10$$

almost surely. For $|t| \le \delta$ and large n,

$$\begin{aligned}&|\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2| + |\hat{\alpha}_n(t) - 1|^2 - t^2/2 \\ &\quad\le \tfrac{1}{8}t^2 + \tfrac{1}{2}t^2 \cdot \tfrac{1}{2}\left(\tfrac{1}{2}\right)^2 - t^2/2 \le -t^2/4.\end{aligned}$$

In this case,

$$|\hat{r}_{n2}(t)| \le C_2\left\{n^2|\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2|^2 + nt^4 + n^2t^8\right\}e^{-nt^2/4}. \tag{4.2}$$

Let $\hat{F}_n$ denote the distribution function corresponding to characteristic function $\hat{\alpha}_n$. Then

$$\begin{aligned}\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2 &= \int\left\{e^{itx} - 1 - itx - \tfrac{1}{2}(itx)^2\right\}d\hat{F}_n(x) \\ &= \tfrac{1}{6}(it)^3\int x^3\,d\hat{F}_n(x) + \hat{r}_{n3}(t),\end{aligned}$$

where $|\hat{r}_{n3}(t)| \le t^4\int x^4\,d\hat{F}_n(x)$. Therefore,

$$|\hat{\alpha}_n(t) - 1 + \tfrac{1}{2}t^2| \le |t|^3n^{-1}\left|\sum_{j=1}^{n}\hat{\sigma}^{-3}\left(X_j - \overline{X}\right)^3\right| + t^4n^{-1}\sum_{j=1}^{n}\hat{\sigma}^{-4}\left(X_j - \overline{X}\right)^4.$$

We may simplify the two series on the right-hand side by arguing as follows. Since $E(X^2) = 1$ and $E(X) = 0$, then by the law of the iterated logarithm, $|\overline{X}| \le 2(n^{-1}\log\log n)^{1/2}$ for all sufficiently large n (f.a.s.l.n). Also, $\sum_j X_j^2 \le 2n$

and $\hat{\sigma}^{-1} \le 2$ f.a.s.l.n. Therefore,

$$\begin{aligned}\left|\sum_{j=1}^{n} \hat{\sigma}^{-3}\left(X_j - \bar{X}\right)^3\right| &\le 8\left|\sum_{j=1}^{n} X_j^3 - 3\bar{X}\sum_{j=1}^{n} X_j^2 + 2n\bar{X}^3\right| \\ &\le 8\left|\sum_{j=1}^{n} X_j^3\right| + 8 \cdot 3 \cdot 2 \cdot 2(n \log\log n)^{1/2} \\ &\quad + 8 \cdot 2 \cdot 2^3 n\left(n^{-1}\log\log n\right)^{3/2},\end{aligned} \tag{4.3}$$

$$\begin{aligned}\sum_{j=1}^{n} \hat{\sigma}^{-4}\left(X_j - \bar{X}\right)^4 &\le 16\left(\sum_{j=1}^{n} X_j^4 - 4\bar{X}\sum_{j=1}^{n} X_j^3 + 6\bar{X}^2\sum_{j=1}^{n} X_j^2 - 3n\bar{X}^4\right) \\ &\le 16\sum_{j=1}^{n} X_j^4 + \left|\sum_{j=1}^{n} X_j^3\right| + 16 \cdot 2 \cdot 2^2 \cdot 2\left(n^{-1}\log\log n\right)n\end{aligned} \tag{4.4}$$

f.a.s.l.n. Combining (4.2) with the results in this paragraph, we deduce that

$$\begin{aligned}|\hat{r}_{n2}(t)| \le C_3\Bigg\{&t^6(1 + t^2)\left(\sum_{j=1}^{n} X_j^3\right)^2 + t^8\left(\sum_{j=1}^{n} X_j^4\right)^2 \\ &+ t^4 n + t^6 n \log\log n + t^8 n^2\Bigg\} e^{-nt^2/4}.\end{aligned} \tag{4.5}$$

The distribution of S^*, conditional on $\mathscr{X}$, has characteristic function $\hat{\alpha}_n(t/n^{1/2})^n$, which in view of (4.1) and (4.5) satisfies

$$\begin{aligned}&\left|\hat{\alpha}_n\left(t/n^{1/2}\right)^n - \left[1 + n\left\{\hat{\alpha}_n\left(t/n^{1/2}\right) - 1 + \tfrac{1}{2}\left(t^2/n\right)\right\}\right]e^{-t^2/2}\right| \\ &\quad\le C_4\Bigg\{t^6(1 + t^2)\left(n^{-3/2}\sum_{j=1}^{n} X_j^3\right)^2 \\ &\qquad + t^8\left(n^{-2}\sum_{j=1}^{n} X_j^4\right)^2 + t^4(1 + t^4)n^{-1}\Bigg\}e^{-t^2/4}\end{aligned} \tag{4.6}$$

for $|t| \le \delta n^{1/2}$ and all sufficiently large n. It is easy to verify that the function

$$\tilde{L}_n(x) \equiv \sum_{j=1}^{n} \Phi\left\{x - \left(X_j - \bar{X}\right)\left(n^{1/2}\hat{\sigma}\right)^{-1}\right\} - n\Phi(x) - \tfrac{1}{2}\phi'(x)$$

has Fourier–Stieltjes transform $n\{\hat{\alpha}_n(t/n^{1/2}) - 1 + \frac{1}{2}(t^2/n)\}e^{-t^2/2}$. In the non-lattice case we use inequality (4.6) and the smoothing lemma [Petrov (1975), Theorem 2, page 109] to deduce that for any $\lambda \geq \delta$,

$$\sup_{-\infty < x < \infty} |P(S^* \leq x|\mathscr{X}) - \Phi(x) - \tilde{L}_n(x)|$$
$$\leq C_5\left\{\left(n^{-3/2}\sum_{j=1}^{n} X_j^3\right)^2 + \left(n^{-2}\sum_{j=1}^{n} X_j^4\right)^2 + n^{-1} + \int_{n^{1/2}\delta}^{n^{1/2}\lambda} t^{-1}|\hat{\alpha}_n(t/n^{1/2})|^n\,dt + n^{-1/2}\lambda^{-1}\right\}$$

f.a.s.l.n. Arguing as in Singh [(1981), page 1190] or as in the proof of result (4.3) of Hall [(1982), page 162], we deduce the existence of a random variable $\lambda_n \geq \delta$ such that $\lambda_n \to \infty$ almost surely and

$$n^{1/2}\int_{n^{1/2}\delta}^{n^{1/2}\lambda_n} |\hat{\alpha}_n(t/n^{1/2})|^n\,dt \to 0$$

almost surely. (This result follows from the fact that $|\hat{\alpha}_n - \alpha| \to 0$ almost surely, uniformly on compacts, and $|\alpha| \leq C < 1$, uniformly on compact intervals not including the origin.) Thus for any $\varepsilon > 0$,

$$\sup_{-\infty < x < \infty} |P(S^* \leq x|\mathscr{X}) - \Phi(x) - \tilde{L}_n(x)|$$
$$\leq C_5\left\{\left(n^{-3/2}\sum_{j=1}^{n} X_j^3\right)^2 + \left(n^{-2}\sum_{j=1}^{n} X_j^4\right)^2\right\} + \varepsilon n^{-1/2} \tag{4.7}$$

f.a.s.l.n.

In the lattice case, it follows as in Hall [(1982), page 170] that the function $\hat{R}_n$ has Fourier–Stieltjes transform

$$\hat{\rho}_n(t) \equiv -(2\pi n^{1/2}\hat{\sigma})^{-1}bt \sum_{\substack{-\infty < j < \infty \\ j \neq 0}} j^{-1}\exp\left[-\tfrac{1}{2}\{t + 2n^{1/2}\pi\hat{\sigma}(j/b)\}^2\right].$$

Arguing as in Singh [(1981), page 1181] or using the argument leading to result (4.23) of Hall [(1982), page 172], we may prove that

$$\int_{n^{1/2}\delta}^{n} t^{-1}|\hat{\rho}_n(t)|\,dt = O(n^{-1}),$$

$$\int_{n^{1/2}\delta}^{n} t^{-1}|\hat{\alpha}_n(t/n^{1/2})^n - \left[1 + n\{\hat{\alpha}_n(t/n^{1/2}) - 1 + \tfrac{1}{2}(t^2/n)\}\right]e^{-t^2/2} - \hat{\rho}_n(t)|\,dt$$
$$= O(n^{-1})$$

almost surely. It then follows from (4.6) via the smoothing lemma that

$$\sup_{-\infty < x < \infty} |P(S^* \le x|\mathscr{X}) - \Phi(x) - \tilde{L}_n(x) - \hat{R}_n(x)|$$

$$\le C_6\left\{\left(n^{-3/2} \sum_{j=1}^{n} X_j^3\right)^2 + \left(n^{-2} \sum_{j=1}^{n} X_j^4\right)^2 + n^{-1}\right\} \tag{4.8}$$

f.a.s.l.n.

The theorem follows from (4.7) and (4.8), provided we show that

$$\sup_{-\infty < x < \infty} |\hat{L}_n(x) - \tilde{L}_n(x)| = o\left(n^{-3/2}\left|\sum_{j=1}^{n} X_j^3\right| + n^{-2} \sum_{j=1}^{n} X_j^4 + n^{-1/2}\right) \tag{4.9}$$

almost surely. Now,

$$\tilde{L}_n(x) = -(6\hat{\sigma}^3)^{-1} n^{-3/2}\left\{\sum_{j=1}^{n} \left(X_j - \overline{X}\right)^3\right\}\phi''(x)$$

$$+ (6\hat{\sigma}^4)^{-1} n^{-2} \sum_{j=1}^{n} \left[\left(X_j - \overline{X}\right)^4 \int_0^1 \phi'''\left\{x - t\left(X_j - \overline{X}\right)(n^{1/2}\hat{\sigma})^{-1}\right\}\right.$$

$$\left. \times (1 - t)^3\, dt\right],$$

$$\hat{L}_n(x) = -\tfrac{1}{6} n^{-3/2}\left(\sum_{j=1}^{n} X_j^3\right)\phi''(x)$$

$$+ \tfrac{1}{6} n^{-2} \sum_{j=1}^{n} \left\{X_j^4 \int_0^1 \phi'''\left(x - tX_j n^{-1/2}\right)(1 - t)^3\, dt\right\}.$$

Expanding $\Sigma_j(X_j - \overline{X})^3$ and $\Sigma_j(X_j - \overline{X})^4$, and noting that $\hat{\sigma} \to 1$ almost surely, we see that

$$D_1 \equiv \left|\hat{\sigma}^{-3} n^{-3/2} \sum_{j=1}^{n} \left(X_j - \overline{X}\right)^3 - n^{-3/2} \sum_{j=1}^{n} X_j^3\right|$$

$$= o\left(n^{-3/2}\left|\sum_{j=1}^{n} X_j^3\right| + n^{-1/2}\right),$$

$$D_2 \equiv n^{-2} \sum_{j=1}^{n} \left|\hat{\sigma}^{-4}\left(X_j - \overline{X}\right)^4 - X_j^4\right|$$

$$= o(1) n^{-2} \sum_{j=1}^{n} X_j^4 + O(1) n^{-2} \sum_{j=1}^{n} |\left(X_j - \overline{X}\right)^4 - X_j^4|.$$

Since $E(X^2) < \infty$, then $P(|X_n| > n^{1/2} \text{ i.o.}) = 0$, and from this it follows that with probability 1, the event $\mathscr{E}_n \equiv \{|X_j| \le n^{1/2} \text{ for } 1 \le j \le n\}$ occurs f.a.s.l.n. If

$\mathscr{E}_n$ occurs, then

$$n^{-2} \sum_{j=1}^{n} |(X_j - \bar{X})^4 - X_j^4|$$

$$\le n^{-2} \sum_{j=1}^{n} \left(4|\bar{X}X_j^3| + 6\bar{X}^2X_j^2 + 4|\bar{X}^3X_j| + \bar{X}^4\right)$$

$$\le \left(4n^{-1/2}|\bar{X}| + 6n^{-1}\bar{X}^2\right)n^{-1} \sum_{j=1}^{n} X_j^2 + 4|\bar{X}|^3 n^{-2} \sum_{j=1}^{n} |X_j| + n^{-1}\bar{X}^4$$

$$= o(n^{-1/2})$$

almost surely. Therefore,

$$D_2 = o\left(n^{-2} \sum_{j=1}^{n} X_j^4 + n^{-1/2}\right).$$

Also, if $\mathscr{E}_n$ occurs then $n^{-5/2}\sum_j |X_j|^5 \le n^{-2}\sum_j X_j^4$, so that

$$D_3 \equiv n^{-2} \sum_{j=1}^{n} \left[X_j^4 \sup_{-\infty < x < \infty} \int_0^1 |\phi'''\left\{x - t(X_j - \bar{X})(n^{1/2}\hat{\sigma})^{-1}\right\} - \phi'''\left(x - tX_j n^{-1/2}\right)|(1-t)^3\, dt \right]$$

$$\le C_7 n^{-2} \sum_{j=1}^{n} X_j^4\left\{\left(|X_j| + |\bar{X}|\right)|\hat{\sigma}^{-1} - 1|n^{-1/2} + |\bar{X}|n^{-1/2}\right\}$$

$$= C_7|\hat{\sigma}^{-1} - 1|n^{-5/2} \sum_{j=1}^{n} |X_j|^5 + o\left(n^{-2} \sum_{j=1}^{n} X_j^4\right)$$

$$= o\left(n^{-2} \sum_{j=1}^{n} X_j^4\right).$$

Combining these results, and noting that

$$6 \sup_{-\infty < x < \infty} |\hat{L}_n(x) - \tilde{L}_n(x)| \le D_1 \sup_{-\infty < x < \infty} |\phi''(x)| + D_2 \sup_{-\infty < x < \infty} |\phi'''(x)| + D_3,$$

we obtain (4.9). □

Proof of Theorem 2.1. Part (i) of the theorem is trivial, since

$$\left|\hat{L}_n(x) - \tfrac{1}{6}(-n^{-1/2})^3\left(\sum_{j=1}^{n} X_j^3\right)\phi''(x)\right| \le \tfrac{1}{24}(\sup|\phi'''|)n^{-2} \sum_{j=1}^{n} X_j^4. \tag{4.10}$$

To obtain part (ii), notice that the function $f(u) \equiv \Phi(1+u) - \Phi(1) - u\phi(1) - \frac{1}{2}u^2\phi'(1)$ is positive for $u \ne 0$ and satisfies $f(u) \ge C_3 \min(u^2, u^4)$ for

$-\infty < u < \infty$. Therefore,

$$|\hat{L}_n(1)| \geq C_3\left\{n^{-1} \sum_{j=1}^{n} X_j^2 I\big(|X_j| > n^{1/2}\big) + n^{-2} \sum_{j=1}^{n} X_j^4 I\big(|X_j| \leq n^{1/2}\big)\right\}.$$

As shown during the proof of Theorem 2.2, we have $\{|X_j| \leq n^{1/2}$ for $1 \leq j \leq n\}$ f.a.s.l.n, so that $\sum_j X_j^2 I(|X_j| > n^{1/2}) = 0$ f.a.s.l.n. Therefore, $|\hat{L}_n(1)| \geq C_3 n^{-2}\sum_j X_j^4$ f.a.s.l.n. An identical lower bound applies to $|\hat{L}_n(-1)|$. Also, by (4.10),

$$|\hat{L}_n(x_2)| \geq \tfrac{1}{6}|\phi''(x_2)| n^{-3/2}\left|\sum_{j=1}^{n} X_j^3\right| - C_4 n^{-2} \sum_{j=1}^{n} X_j^4,$$

where $C_4 \equiv \frac{1}{24}(\sup|\phi'''|)$. Hence f.a.s.l.n,

$$\sup_{x \in \{x_1, x_2\}} |\hat{L}_n(x)| \geq C_5 \hat{\delta}_n \quad \text{where } C_5 \equiv C_3 \min\left\{1, (C_3 + C_4)^{-1} \tfrac{1}{6}|\phi''(x_2)|\right\}. \quad \square$$

PROOF OF LEMMA 3.1. That $T_n/c_n \to 0$ implies $\sum P(|Y| > c_n) < \infty$ follows from Petrov [(1975), Lemma 14, page 273]. Suppose $\sum P(|Y| > c_n) < \infty$. Since $b_n \equiv n^{-1}c_n$ satisfies (3.4), then either $n^{-1}c_n$ is bounded away from zero and infinity or $n^{-1}c_n \to +\infty$. In the former case, $\sum P(|Y| > c_n) < \infty$ is equivalent to $E(|Y|) < \infty$, which implies $T_n/c_n \to 0$ [since by assumption, $E(Y) = 0$]. In the latter case, the result $T_n/c_n \to 0$ will follow via Petrov [(1975), Theorem 16, page 274] if we prove that

$$\sum_{n=1}^{\infty} c_n^{-2} E\{Y^2 I(|Y| \leq c_n)\} < \infty \quad \text{and} \quad n c_n^{-1} E\{|Y| I(|Y| \leq c_n)\} \to 0.$$

The first of these results may be established as in Petrov [(1975), page 275], while the second follows from the fact that for all $n \geq m$,

$$\begin{aligned} n c_n^{-1} E\{|Y| I(|Y| \leq c_n)\} &\leq n c_n^{-1} c_m + n c_n^{-1} \sum_{j=m+1}^{n} j\big(j^{-1} c_j\big) P\big(c_{j-1} < |Y| \leq c_j\big) \\ &\leq o(1) + C \sum_{j=m+1}^{\infty} j P\big(c_{j-1} < |Y| \leq c_j\big) \\ &= o(1) + m P(|Y| > c_m) + C \sum_{j=m}^{\infty} P\big(|Y| > c_j\big), \end{aligned}$$

where C is as in (3.4). $\square$

Acknowledgment. I am grateful to a referee for a carefully and thoroughly compiled list of places where presentation could be improved. The paper has benefited from those points.

REFERENCES

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

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

BABU, G. J. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003.

BABU, G. J. and SINGH, K. (1984). Asymptotic representations related to jackknifing and bootstrapping L-statistics. *Sankhyā Ser. A* **46** 195–206.

BERAN, R. (1984). Bootstrap methods in statistics. *Jahresber. Deutsch. Math.-Verein.* **86** 14–30.

BICKEL, P. J. and FREEDMAN, D. A. (1980). On Edgeworth expansions for the bootstrap. Unpublished.

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

CHUNG, K. L. (1974). *A Course in Probability Theory.* Academic, New York.

EFRON, B. (1981). Nonparametric standard errors and confidence intervals (with discussion). *Canad. J. Statist.* **9** 139–172.

EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statist. Sci.* **1** 54–77.

FELLER, W. (1946). A limit theorem for random variables with infinite moments. *Amer. J. Math.* **68** 257–262.

GNEDENKO, B. V. and KOLMOGOROV, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, Mass.

HALL, P. (1982). *Rates of Convergence in the Central Limit Theorem.* Pitman, London.

HÖGLUND, T. (1970). On the convergence of convolutions of distributions with regularly varying tails. *Z. Wahrsch. verw. Gebiete* **15** 263–272.

KESTEN, H. (1972). Sums of independent random variables without moment conditions. *Ann. Math. Statist.* **43** 701–732.

KLASS, M. and TEICHER, H. (1977). Iterated logarithm laws for asymmetric random variables barely with or without finite mean. *Ann. Probab.* **5** 861–874.

PETROV, V. V. (1975). *Sums of Independent Random Variables.* Springer, Berlin.

SENETA, E. (1976). *Regularly Varying Functions. Lecture Notes in Math.* **508**. Springer, New York.

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

STOUT, W. F. (1974). *Almost Sure Convergence.* Academic, New York.

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