*The Annals of Probability*
1990, Vol. 18, No. 3, 1342–1360

# ASYMPTOTIC PROPERTIES OF THE BOOTSTRAP FOR HEAVY-TAILED DISTRIBUTIONS[1]

BY PETER HALL

*Brown University*

We establish necessary and sufficient conditions for convergence of the distribution function of a bootstrapped mean, suitably normalized. It turns out that for convergence to occur, the sampling distribution must either be in the domain of attraction of the normal distribution or have slowly varying tails. In the first case the limit is normal; in the latter, Poisson. Between these two extremes of light tails and extremely heavy tails, the bootstrap distribution function of the mean does not converge in probability to a nondegenerate limit. However, it may converge in distribution. We show that when there is no convergence in probability, a small number of extreme sample values determine behaviour of the bootstrap distribution function. This result is developed and used to interpret recent work of Athreya.

**1. Introduction.** Properties of the bootstrap for heavy-tailed distributions have recently been the subject of attention in statistical literature; see Athreya [2, 3]. It has been pointed out that the bootstrap is not consistent for estimating the distribution of the mean when the parent population is from the domain of attraction of a nonnormal stable law. However, there has been no precise description of circumstances where the bootstrap estimate of a distribution function even converges, in particular where it is consistent for the limiting distribution function of the mean. The aim of the present paper is to provide such an account. We establish, among other things, the following results.

1. The bootstrap distribution function of the mean, suitably normalized, converges in probability to some fixed nondegenerate distribution function if and only if either (a) the sampling distribution is from the domain of attraction of the normal law or (b) the sampling distribution has slowly varying tails and one of the two tails completely dominates the other. Thus, convergence occurs only in the case of light tails or extremely heavy tails. For sampling distributions between these two extremes the bootstrap distribution function does not converge in probability. [Theorem 2.1.]
2. In case (a), the limiting distribution is normal. In case (b) it is Poisson with unit mean. [Theorem 2.1.] Only case (a) is statistically interesting, since in case (b) the limit of the bootstrap distribution function does not reflect statistical properties of the distribution of the mean, which is not asymptotically Poisson-distributed. [Remark 2.3.] Therefore, the bootstrap is (weakly)

Received July 1988; revised June 1989.

[1]On leave from Australian National University.

*AMS* 1980 *subject classifications*. Primary 60F05, 60G50; secondary 62G05.

*Key words and phrases*. Bootstrap, central limit theorem, domain of attraction, heavy tail, normal distribution, stable law.

consistent if and only if the sampling distribution is from the domain of attraction of the normal law.

3. There are analogous results in the case of almost sure convergence. In particular, the bootstrap distribution of the mean converges almost surely to normality if and only if the sampling distribution has finite variance. [Proposition 2.1 and Remark 2.7.] This condition is strictly stronger than the assumption that it be in the domain of attraction of the normal distribution.
4. The reason the bootstrap fails for intermediate distributions, such as those in domains of attraction of stable laws, is that in these circumstances the bootstrap does not correctly model the way in which extreme terms influence the distribution of a sum. We present a general result, available more widely than for domains of attraction of stable laws, which describes the way in which the bootstrap distribution of the mean is determined by a small number of extreme sample values. [Theorem 2.2 and Remark 2.9.]
5. The bootstrap distribution function does not converge in probability in such cases because extreme values of the sample do not converge in probability, even when suitably normalized. [Remark 2.10.] However, those extreme values may converge in distribution, and then the bootstrap distribution function converges in distribution. Indeed, it has a weak limit when viewed as a stochastic process. [Remark 2.12.]
6. These results enable us to give new interpretations of the work of Athreya [2] on properties of the bootstrap when the sampling distribution is from the domain of attraction of a stable law. [Remark 2.13.]

So as to present these conclusions economically and succinctly we have chosen to state our basic results in two main theorems and to discuss their implications and generalizations in remarks following those theorems. Some of the remarks comprise corollaries to the theorems; others describe related results which may be derived with only slight modifications of our proofs.

The following notation is used. Let $X_1, X_2, \dots$ be independent random variables with nondegenerate distribution function $F$, write $X$ for a generic $X_i$ and let $\mathscr{X} = \{X_1, \dots, X_n\}$ be the sample of the first $n$ $X_i$'s. Write $\mathscr{X}^* = \{X_1^*, \dots, X_n^*\}$ for a resample drawn randomly (with replacement) from $\mathscr{X}$ and let $\overline{X} = n^{-1}\Sigma_{i \le n} X_i$ and $\overline{X}^* = n^{-1}\Sigma_{i \le n} X_i^*$ denote the respective sample means. Put $\hat{\sigma}^2 = n^{-1}\Sigma_{i \le n}(X_i - \overline{X})^2$. The bootstrap distribution function of the mean, suitably normalized, is

$$P\{(\overline{X}^* - A_n)/B_n \le x | \mathscr{X}\}, \qquad -\infty < x < \infty,$$

where $A_n$ and $B_n$ are (arbitrary) measurable functions of elements of $\mathscr{X}$. We say that $A_n$ and $B_n$ are $\mathscr{X}$-measurable. Convergence in probability is denoted by $\to_p$. We intend the statements "$X$ is in the domain of attraction of ..." and "$F$ is in the domain of attraction of ..." to mean the same thing.

**2. Results.** We begin by characterizing the class of situations where there exists a nonrandom, nondegenerate distribution function $G$ such that, for some choice of $\mathscr{X}$-measurable random variables $A_n$ and $B_n$ and at continuity points $x$ of $G$,

$$P\{(\bar{X}^* - A_n)/B_n \le x | \mathscr{X}\} \to_p G(x). \tag{2.1}$$

The zero-one law guarantees that no interesting circumstances are excluded by our insistence that $G$ be nonrandom.

THEOREM 2.1. *There exist $\mathscr{X}$-measurable random variables $A_n$ and $B_n$ such that* (2.1) *holds for a nonrandom, nondegenerate $G$, if and only if either*

$$1 - F \text{ is slowly varying at } +\infty \text{ and } P(X < -x)/P(|X| > x) \to 0 \text{ as } x \to \infty \tag{2.2}$$

*or*

$$F \text{ is slowly varying at } -\infty \text{ and } P(X > x)/P(|X| > x) \to 0 \text{ as } x \to \infty \tag{2.3}$$

*or*

$$F \text{ is in the domain of attraction of the normal distribution.} \tag{2.4}$$

*In the circumstance of* (2.2), *if we take $A_n = \bar{X}$ and $nB_n = \max_{i \le n}|X_i|$, then $G(x) = P(N - 1 \le x)$ where $N$ is Poisson with unit mean. In the circumstance of* (2.3), *if we take $A_n = \bar{X}$ and $nB_n = \max_{i \le n}|X_i|$, then $G(x) = P(1 - N \le x)$, for the same $N$ as before. In the circumstance of* (2.4), *if we take $A_n = \bar{X}$ and $nB_n = \{\Sigma_{i \le n}(X_i - \bar{X})^2\}^{1/2}$, then $G$ is the standard normal distribution function.*

REMARK 2.1. The *only* possible limit distributions $G$, modulo changes of scale and location, are Poisson with unit mean, negative Poisson with unit mean, and normal, as described in Theorem 2.1. This follows from the convergence of types theorem. Specifically, suppose that there exist $\mathscr{X}$-measurable sequences $A_n$, $B_n$, $A_n^\dagger$ and $B_n^\dagger$, and nondegenerate distribution functions $G$ and $G^\dagger$, such that

$$P\{(\bar{X}^* - A_n)/B_n \le x | \mathscr{X}\} \to_p G(x), \quad P\{(\bar{X}^* - A_n^\dagger)/B_n^\dagger \le y | \mathscr{X}\} \to_p G^\dagger(y)$$

at continuity points. Choose a subsequence of $n$ values along which the first convergence in probability is almost sure. Then choose a subsubsequence along which both convergences are almost sure. Now apply the classical convergence of types theorem [10, page 40] for any of the probability-1 realizations and along the subsubsequence, to deduce that $G$ and $G^\dagger$ are scale and location changes of one another. This subsequence argument is used implicitly in several steps of our proof of Theorem 2.1.

REMARK 2.2. The variable $X$, or distribution function $F$, is in the domain of attraction of the normal distribution if and only if $E\{X^2 I(|X| \le x)\}$ is a slowly varying function of $x$, as $x \to \infty$. This implies $E|X|^{2-\varepsilon} < \infty$ for each $0 < \varepsilon < 2$ [12, pages 83, 84]. In particular it implies finite mean.

REMARK 2.3. Distributions $F$ satisfying (2.2) or (2.3) have no finite moments: $E|X|^{\varepsilon} = \infty$ for all $\varepsilon > 0$. They are so heavy-tailed that sums are entirely dominated by extreme terms, in the sense that

$$\left( \sum_{i=1}^{n} |X_i| \right) \Big/ \left( \max_{1 \le i \le n} |X_i| \right) \to_p 1$$

as $n \to \infty$ [8, 9, Section 4.5]. In fact, this property explains how the limit laws in cases (2.2) and (2.3) come about. There the mean $\overline{X}^*$ is entirely dominated by appearances of $\max_{i \le n} X_i$ [in case (2.2)] or $\min_{i \le n} X_i$ [in case (2.3)] in $\mathscr{X}$, and the asymptotic Poisson distribution is simply the limit of the number of times that this extreme value appears in $\mathscr{X}^*$. Therefore, when (2.2) or (2.3) holds, the limit theorem (2.1) does no more than record the limit of the number of times a certain sample value appears in the resample. The limit theorem does not reflect statistical properties of the distribution of $\overline{X}$, which does not have an asymptotic Poisson distribution [10, page 132]. Therefore, the only statistically interesting case is that of the normal law, in circumstance (2.4).

Remarks 2.4–2.8 record consequences of Theorem 2.1 and also related results which may be proved in like manner.

REMARK 2.4. Let $X$ be in the domain of attraction of the normal distribution and assume (without loss of generality) that $E(X) = 0$. Take $b_n$ to be any sequence of positive numbers satisfying

$$b_n^{-2} E\{X^2 I(|X| \le n^{1/2} b_n)\} \to 1 \tag{2.5}$$

as $n \to \infty$. Then we may take $A_n = \overline{X}$ and $B_n = b_n / n^{1/2}$ in (2.1), obtaining a standard normal limit. This follows from Theorem 2.1 on noting that for the $B_n$ given in that theorem and $b_n$ given by (2.5),

$$\begin{aligned} nB_n^2/b_n^2 &= (nb_n^2)^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \\ &= (nb_n^2)^{-1} \sum_{i=1}^{n} X_i^2 + o_p(1) \to_p 1 \end{aligned}$$

[10, Section 28]. If $E(X^2) = \sigma^2 < \infty$, then $b_n \to \sigma$ and we may take $A_n = \overline{X}$ and $B_n = \sigma / n^{1/2}$.

REMARK 2.5. Even in the case $E(X^2) < \infty$ and $E(X) = 0$ it is not possible to take $A_n = 0$ in (2.1). This follows from the fact that $n^{1/2}(\overline{X} - EX)$ does not converge to zero in probability.

REMARK 2.6. In view of the result described in Remark 2.4, the following is true: The limit theorem (2.1), with normal G and a *nonrandom* choice of $B_n$, holds if and only if it is available with a *random* choice of $B_n$.

REMARK 2.7. A minor modification of part of our proof of Theorem 2.1 allows us to prove the following. (Let $\Phi$ denote the standard normal distribution function.) There exist $\mathscr{X}$-measurable random variables $A_n$, $B_n$ such that

$$P\{(\bar{X}^* - A_n)/B_n \le x|\mathscr{X}\} \to \Phi(x), \qquad -\infty < x < \infty, \tag{2.6}$$

if and only if

$$\Big(\max_{1 \le i \le n} X_i^2\Big) \Big/ \Big(\sum_{i=1}^{n} X_i^2\Big) \to 0. \tag{2.7}$$

In this result, $\to$ may be interpreted as convergence in probability in both cases or as almost sure convergence in both places. We prove the equivalence of (2.6) and (2.7) when $\to$ is $\to_p$ ; to obtain equivalence in the case of almost sure convergence, simply interpret all convergences in that part of our proof in a strong rather than a weak sense. Of course, (2.7) with convergence in probability is equivalent to $X$ being in the domain of attraction of the normal law (see Theorem 2.1). It may be shown that (2.7) with almost sure convergence is equivalent to $E(X^2) < \infty$ (see Lemma 3.3 in Section 3) and the latter is equivalent to $X$ being in the normal domain of attraction of the normal law [12, page 92]. Thus, we have the following result.

PROPOSITION 2.1. *There exist $\mathscr{X}$-measurable random variables $A_n$, $B_n$ such that* (2.6) *holds with convergence in probability* [*respectively, almost sure convergence*] *if and only if $X$ is in the domain of attraction* [*respectively, normal domain of attraction*] *of the normal law, or equivalently, if and only if $E\{X^2 I(|X| \le x)\}$ is slowly varying* [*respectively,* $E(X^2) < \infty$]. *In both cases we may take $A_n = \bar{X}$ and $nB_n = \{\sum_{i \le n}(X_i - \bar{X})^2\}^{1/2}$.*

REMARK 2.8. A necessary and sufficient condition for (2.6) to hold for a *nonrandom* choice of $B_n$, when convergence in (2.6) is almost sure, is $E(X^2) < \infty$. In that circumstance we may take $A_n = \bar{X}$ and $B_n = n^{-1/2}\sigma$, where $\sigma^2 = \mathrm{var}(X)$.

One conclusion which may be drawn from Theorem 2.1 is that the bootstrap fails unless the sampling distribution has either light tails (i.e., is in the domain of attraction of the normal law) or extremely heavy tails (i.e., has slowly varying tails). In the remainder of this section we explain what goes on between these two extremes.

A necessary and sufficient condition for $X$ to be in the domain of attraction of the normal law is

$$E\{X^2 I(|X| \le x)\}/\{x^2 P(|X| > x)\} \to \infty$$

as $x \to \infty$ [12, page 84]. In the present investigation we are bound to assume that this condition fails and so we suppose that

$$E\{X^2 I(|X| \le x)\}/\{x^2 P(|X| > x)\} \le C_1 < \infty, \quad \text{all } x \ge 1 \tag{2.8}$$

(which is a little stronger than the failure of the previous condition). We further assume that

$$\lim_{u \to 0} u^2 \limsup_{x \to \infty} P(|X| > ux)/P(|X| > x) = 0. \tag{2.9}$$

These conditions are weaker than the assumption that $X$ be in the domain of attraction of a nonnormal stable law, which requires among other assumptions that for some $0 < \alpha < 2$ and each $u > 0$,

$$P(|X| > ux)/P(|X| > x) \to u^{-\alpha} \tag{2.10}$$

as $x \to \infty$.

When the sampling distribution satisfies (2.8) and (2.9), a natural scaling sequence for a sum of independent random variables is a sequence $c_n$ such that $nP(|X| > c_n) \to 1$. For example, if $X$ is in the domain of attraction of a stable law and $c_n$ has this property, there exist constants $a_n$ such that $nc_n^{-1}(\bar{X} - a_n)$ converges in distribution to that stable law. We ask only that $nP(|X| > c_n)$ be bounded.

We shall prove that for distributions which lie between the two extremes of light tails and extremely heavy tails, the bootstrap fails because it does not correctly model the way in which extreme summands contribute to a sum. In more detail, let $\{X_{\langle n,i \rangle}, 1 \le i \le n\}$ denote a rearrangement of the sample $\mathscr{X}$ such that $|X_{\langle n,1 \rangle}| \ge |X_{\langle n,2 \rangle}| \ge \cdots$. Let $N_i^*$ denote the number of times that $X_{\langle n,i \rangle}$ appears in the resample $\mathscr{X}^*$. Then

$$c_n^{-1} \sum_{i=1}^{r} (N_i^* - 1) X_{\langle n,i \rangle}$$

equals the total contribution of $X_{\langle n,1 \rangle}, \dots, X_{\langle n,r \rangle}$ to the normalized sum $nc_n^{-1}(\bar{X}^* - \bar{X})$. Subtracting this contribution away from the normalized sum we obtain the quantity

$$\Delta_{nr} = nc_n^{-1}(\bar{X}^* - \bar{X}) - c_n^{-1} \sum_{i=1}^{r} (N_i^* - 1) X_{\langle n,i \rangle}.$$

Our next result shows that $\Delta_{nr}$ may be rendered arbitrarily small uniformly in $n$, simply by choosing $r$ large.

THEOREM 2.2. *Assume conditions* (2.8) *and* (2.9) *and that* $nP(|X| > c_n)$ *is bounded. Then for each* $\varepsilon > 0$,

$$\lim_{r \to \infty} \limsup_{n \to \infty} P\{P(|\Delta_{nr}| > \varepsilon | \mathscr{X}) > \varepsilon\} = 0.$$

REMARK 2.9. Theorem 2.2 may be interpreted in the following way. Given $\varepsilon > 0$, we may choose a *fixed* $r \ge 1$ so large that for all $n \ge 1$,

$$P\{P(|\Delta_{nr}| > \varepsilon | \mathscr{X}) > \varepsilon\} < \varepsilon.$$

Then except for an error smaller than $\varepsilon$, the distribution of $nc_n^{-1}(\bar{X}^* - \bar{X})$

conditional on $\mathscr{X}$ is equivalent to that of

$$(2.11)\qquad c_n^{-1} \sum_{i=1}^{r} (N_i^* - 1) X_{\langle n, i \rangle}.$$

Therefore, the size of $\bar{X}^* - \bar{X}$ is essentially determined by a relatively small and fixed number of extreme sample values. It now becomes clear that asymptotic properties of $\bar{X}^* - \bar{X}$ are readily described in terms of extreme value theory. The ensuing remarks will develop this argument in several different directions.

For the remainder of this section we assume that $X$ is in the domain of attraction of a nonnormal stable law.

REMARK 2.10. Suppose $X$ is in the domain of attraction of a stable law with exponent $\alpha$. That is, (2.10) holds and for some $0 \le p \le 1$,

$$(2.12)\qquad P(X > x)/P(|X| > x) \to p$$

as $x \to \infty$. Define $c_n$ so that

$$(2.13)\qquad nP(|X| > c_n) \to 1$$

as $n \to \infty$. The variables $N_i^*$ appearing in (2.11) have a multinomial distribution with probabilities $p_i = n^{-1}$ and have a joint distribution not depending on $\mathscr{X}$. Therefore,

$$P(N_i^* \le x_i, 1 \le i \le r) \to \prod_{i=1}^{r} P(N \le x_i),$$

where $N$ is Poisson with unit mean. However, the random variables $c_n^{-1} X_{\langle n, i \rangle}$ in (2.11) do not converge in probability. This means that the distribution function

$$P\left\{ c_n^{-1} \sum_{i=1}^{r} (N_i^* - 1) X_{\langle n, i \rangle} \le x \,\middle|\, \mathscr{X} \right\}$$

does not converge in probability as $n \to \infty$. It is not difficult to prove from these observations that the distribution function

$$G_n(x) = P\{ n c_n^{-1} (\bar{X}^* - \bar{X}) \le x \mid \mathscr{X} \}$$

does not converge in probability either; indeed, that follows from Theorem 2.1. However, $G_n$ and related functions do converge weakly, as we shall show in Remarks 2.12 and 2.13.

REMARK 2.11. If the distribution of $X$ satisfies (2.10) and (2.12) and $c_n$ satisfies (2.13), then the variables $c_n^{-1} X_{\langle n, i \rangle}$ converge weakly. To appreciate the form of the limit, recall from extreme value theory [9, Section 2.8; 14, 16] that for each $r \ge 1$,

$$(2.14)\qquad \left( c_n^{-1} |X_{\langle n, 1 \rangle}| \dots c_n^{-1} |X_{\langle n, r \rangle}| \right) \to (Y_1 \dots Y_r)$$

in distribution, where $Y_1, \ldots, Y_r$ have a joint distribution with marginals given by

$$P(Y_j \le y) = \sum_{i=0}^{j-1} (i! y^{i\alpha})^{-1} \exp(-y^{-\alpha}), \qquad y > 0. \tag{2.15}$$

Let $S_1, S_2, \ldots$ be independent random variables, independent also of $Y_1, Y_2, \ldots$ and satisfying

$$P(S_i = 1) = p = 1 - P(S_i = -1). \tag{2.16}$$

In view of the balancing condition (2.12) we may deduce from (2.14) that

$$\left(c_n^{-1} X_{\langle n,1\rangle}, \ldots, c_n^{-1} X_{\langle n,r\rangle}\right) \to (S_1 Y_1, \ldots, S_r Y_r) \tag{2.17}$$

in distribution as $n \to \infty$.

In Remarks 2.12–2.14 we use notation from Remark 2.11.

REMARK 2.12. If $Y_i$ has distribution given by (2.15), then $E(Y_i^2) = \Gamma(i - 2\alpha^{-1})/\Gamma(i)$ and so $\sum_{i > 2/\alpha} E(Y_i^2) < \infty$. Using Kolmogorov's extension theorem we may construct an infinite sequence $Y_1, Y_2, \ldots,$ with finite dimensional distributions satisfying (2.14) for each $r \ge 1$ and such that $\sum_{i \ge 1} Y_i^2 < \infty$ almost surely. Let $N_1, N_2, \ldots$ and $S_1, S_2, \ldots$ be independent random variables, independent of $Y_1, Y_2, \ldots,$ each $N_i$ having the Poisson distribution with unit mean and each $S_i$ having the distribution at (2.16). Since

$$E\left[\left\{\sum_{i=1}^{\infty} (N_i - 1) S_i Y_i\right\}^2 \middle| S_i, Y_i, i \ge 1\right] = \sum_{i=1}^{\infty} Y_i^2 < \infty,$$

then the (random) distribution function

$$G(x) = P\left\{\sum_{i=1}^{\infty} (N_i - 1) S_i Y_i \le x \middle| S_i, Y_i, i \ge 1\right\}, \qquad -\infty < x < \infty,$$

is proper and well defined. We know from Remark 2.10 that the (random) distribution function

$$G_n(x) = P\{n c_n^{-1}(\bar{X}^* - \bar{X}) \le x | \mathscr{X}\}$$

does not converge in probability to anything. However, it follows from Theorem 2.2 and result (2.17) that $G_n(\cdot) \to G(\cdot)$ weakly in $D[-\lambda, \lambda]$ for each $\lambda > 0$. In particular,

$$P\{G_n(x) \le y\} \to P\{G(x) \le y\}.$$

for each $x, y$.

REMARK 2.13. The principal difference between the result in our Remark 2.12 and that proved by Athreya [2] is that Athreya chose to standardize by a random variable, rather than the constant $c_n$. However, we may very easily treat Athreya's problem. Assume $p \neq 0$ in (2.12) and put

$$J_n = \inf\{i \geq 1 : X_{\langle n,i\rangle} > 0\}.$$

Athreya's limit theorem was proved for $G_n^\dagger$, defined by

$$G_n^\dagger(x) = P\{nX_{\langle n,J_n\rangle}^{-1}(\bar{X}^* - \bar{X}) \leq x | \mathscr{X}\},$$

instead of for $G_n$. If we define $J = \inf\{i \geq 1: S_i > 0\}$ and put

$$G^\dagger(x) = P\left\{Y_J^{-1} \sum_{i=1}^{\infty} (N_i - 1) S_i Y_i \leq x \,\middle|\, S_i, Y_i, i \geq 1\right\},$$

we see from Theorem 2.2 and result (2.17) that $G_n^\dagger(\cdot) \to G^\dagger(\cdot)$ weakly in $D[-\lambda, \lambda]$ for each $\lambda > 0$. In particular,

$$P\{G_n^\dagger(x) \leq y\} \to P\{G^\dagger(x) \leq y\}$$

for each $x, y$. This is a generalization of Athreya's [2] result to the stochastic process $G_n^\dagger(\cdot)$, the difference in our exposition being that we have related the limiting process $G^\dagger(\cdot)$ to fluctuations of extremes from the sample $\mathscr{X}$.

REMARK 2.14. There exist versions of the above results for convergence of ordinary (as distinct from bootstrap) sums of independent random variables. Assume the distribution of $X$ satisfies (2.10) and (2.12), and the constants $c_n$ satisfy (2.13). Then there exist constants $a_n$ and a distribution function $H$ of a stable law with exponent $\alpha$ such that

$$P\{nc_n^{-1}(\bar{X} - a_n) \leq x\} \to H(x). \tag{2.18}$$

Define

$$\Delta'_{nr} = nc_n^{-1}\bar{X} - nc_n^{-1} \sum_{i=1}^{r} X_{\langle n,i\rangle},$$

which is just the normalized sample mean with contributions of extreme terms removed. There exist constants $a_{nr}$ such that for each $\varepsilon > 0$,

$$\lim_{r\to\infty} \limsup_{n\to\infty} P(|\Delta'_{nr} - a_{nr}| > \varepsilon) = 0, \tag{2.19}$$

an analogue of Theorem 2.2. We know from Remark 2.11 that

$$c_n^{-1} \sum_{i=1}^{r} X_{\langle n,i\rangle} \to \sum_{i=1}^{r} S_i Y_i$$

in distribution as $n \to \infty$. There exist constants $d_r$ such that

$$P\left(\sum_{i=1}^{r} S_i Y_i - d_r \leq x\right) \to H(x), \qquad -\infty < x < \infty, \tag{2.20}$$

as $r \to \infty$, where $H$ is exactly as in (2.18). Results (2.19) and (2.20) are proved in [1, 11]. See [4–7] for related work.

**3. Proofs.** We begin with a lemma on properties of extreme values, related to work of Smid and Stam [15] but without their assumption of continuity. Let $X_{n1} \leq \cdots \leq X_{nn}$ denote the sample $\mathscr{X}$ arranged in order of increasing magnitude.

LEMMA 3.1. *If for some $\varepsilon > 0$ and $k \geq 1$,*

$$P(X_{n,n-k}/X_{n,n-k+1} > 1 - \varepsilon) \to 0,$$

*then $X_{n,n-l}/X_{n,n-l+1} \to_p 0$ for each $l \geq 1$ and $1 - F$ is slowly varying at infinity.*

Taking logs of the order statistic ratio we may derive an equivalent theorem for the difference between two order statistics.

PROOF. Put $x = (1-\varepsilon)^{-1} > 1$. Then $P(X_{n,n-k+1} > xX_{n,n-k}) \to 1$. Let $U_{n1} < \cdots < U_{nn}$ be the order statistics of a random, uniform $(0,1)$ $n$-sample. Write

$$f(u,v) = n!\{(k-1)!(n-k-1)!\}^{-1} u^{n-k-1}(1-v)^{k-1}$$

for the density of $(U,V) = (U_{n,n-k}, U_{n,n-k+1})$. Since $F^{-1}(v) > xF^{-1}(u)$ implies $v > F\{xF^{-1}(u)\}$, then

$$\begin{aligned}
&P(X_{n,n-k+1} > xX_{n,n-k}) \\
&\quad = \int_0^1 \int_0^1 I\{F^{-1}(v) > xF^{-1}(u)\} f(u,v)\, du\, dv \\
&\quad \leq \int_0^1 \int_0^1 I\left[v > F\{xF^{-1}(u)\}\right] f(u,v)\, du\, dv \\
&\quad = n!\{k!(n-k-1)!\}^{-1} \int_0^1 u^{n-k-1}\left[1 - F\{xF^{-1}(u)\}\right]^k du \\
&\quad = \{1 + o(1)\} n^{k+1}(k!)^{-1} \int_0^1 u^{n-k-1}\left[1 - F\{xF^{-1}(u)\}\right]^k du.
\end{aligned}$$

For any $\delta > 0$, the contribution of the integral over $(0, 1-\delta)$ to the integral on the right-hand side equals $O[\exp\{-C(\delta)n\}]$, where $C(\delta) > 0$. Therefore, we may replace the term $u^{n-k-1}$ in the integrand by $u^n$, without affecting the asymptotics:

$$\begin{aligned}
&P(X_{n,n-k+1} > xX_{n,n-k}) \\
&\quad \leq \{1 + o(1)\} n^{k+1}(k!)^{-1} \int_0^1 u^n\left[1 - F\{xF^{-1}(u)\}\right]^k du + o(1).
\end{aligned}$$

Take $u = 1 - n^{-1}v$ in the last-written integral, obtaining

$$\begin{aligned}
&P(X_{n,n-k+1} > xX_{n,n-k}) \\
&\quad \le \{1 + o(1)\}n^k(k!)^{-1}\int_0^n (1 - n^{-1}v)^n\big[1 - F\{xF^{-1}(1 - n^{-1}v)\}\big]^k\,dv + o(1) \\
&\quad \le n^k(k!)^{-1}\int_0^n \big[1 - F\{xF^{-1}(1 - n^{-1}v)\}\big]^k e^{-v}\,dv + o(1) \\
&\quad \le n(k!)^{-1}\int_0^n u^n\big[1 - F\{xF^{-1}(1 - n^{-1}v)\}\big]v^{k-1}e^{-v}\,dv + o(1) \\
&\quad \le 1 + o(1).
\end{aligned}$$

Therefore, $P(X_{n,n-k+1} > xX_{n,n-k}) \to 1$ entails

$$\int_0^n v^k a(v; n, x)e^{-v}\,dv \to \int_0^\infty v^k e^{-v}\,dv,$$

where $a(v, n, x) = nv^{-1}[1 - F\{xF^{-1}(1 - n^{-1}v)\}]$.

Since $a(v, n, x) \le 1$, then for each $u > 0$ and $\varepsilon > 0$,

$$\begin{aligned}
A(u, \varepsilon; n, x) &= \int_u^{u+\varepsilon} v^k a(v, n, x)e^{-v}\,dv \\
&\to \int_u^{u+\varepsilon} v^k e^{-v}\,dv = A(u, \varepsilon).
\end{aligned} \tag{3.1}$$

Now,

$$\begin{aligned}
A(u, \varepsilon; n, x) &\ge \varepsilon u^k n(u + \varepsilon)^{-1}\big[1 - F\{xF^{-1}(1 - n^{-1}u)\}\big]e^{-(u+\varepsilon)} \\
&= (\varepsilon u^k e^{-u})\big\{(1 + \varepsilon u^{-1})^{-1}e^{-\varepsilon}\big\}nu^{-1} \\
&\qquad \times\big[1 - F\{xF^{-1}(1 - n^{-1}u)\}\big], \\
A(u - \varepsilon, \varepsilon; n, x) &\le \varepsilon u^k n(u - \varepsilon)^{-1}\big[1 - F\{xF^{-1}(1 - n^{-1}u)\}\big]e^{-(u-\varepsilon)} \\
&= (\varepsilon u^k e^{-u})\big\{(1 - \varepsilon u^{-1})^{-1}e^{\varepsilon}\big\}nu^{-1}\big[1 - F\{xF^{-1}(1 - n^{-1}u)\}\big].
\end{aligned}$$

Put

$$\begin{aligned}
f_1(x) &= \liminf_{v\to\infty} v\big[1 - F\{xF^{-1}(1 - v^{-1})\}\big], \\
f_2(x) &= \limsup_{v\to\infty} v\big[1 - F\{xF^{-1}(1 - v^{-1})\}\big].
\end{aligned}$$

Then by the results from (3.1) down,

$$\begin{aligned}
A(u, \varepsilon)/(\varepsilon u^k e^{-u}) &\ge \big\{(1 + \varepsilon u^{-1})^{-1}e^{-\varepsilon}\big\}f_2(x), \\
A(u - \varepsilon, \varepsilon)/(\varepsilon u^k e^{-u}) &\le \big\{(1 - \varepsilon u^{-1})^{-1}e^{-\varepsilon}\big\}f_1(x).
\end{aligned}$$

Fix $u > 0$ and let $\varepsilon \to 0$ in these inequalities, deducing that $f_2(x) \le 1 \le f_1(x)$.

Therefore, $f_1(x) = f_2(x) = 1$. We claim that this implies

$$\{1 - F(xy)\}/\{1 - F(y)\} \to 1 \tag{3.2}$$

as $y \to \infty$. Once this result holds for some $x > 1$, it holds for all $x > 1$, since $F$ is increasing and

$$\{1 - F(x^m y)\}/\{1 - F(y)\} = \prod_{i=1}^{m} \left[\{1 - F(x^i y)\}/\{1 - F(x^{i-1} y)\}\right].$$

Therefore, $1 - F$ is slowly varying. It is now readily proved that for each $k \geq 1$, $X_{n,n-k+1}/X_{n,n-k} \to_p 0$.

We shall conclude the proof of Lemma 3.1 by showing that (3.2) follows from the fact that $f_1(x) = f_2(x) = 1$, that is, from

$$\lim_{v \to \infty} v\left[1 - F\{xF^{-1}(1 - v^{-1})\}\right] = 1.$$

Since $1 - F$ is right continuous, then the above result implies

$$\{1 - F(xt)\}/\{1 - F(t)\} \to 1 \quad \text{as } t \to \infty \text{ through supp } F. \tag{3.3}$$

Fix $\delta > 0$. Given $s$, choose $y = y(s) \geq 1 - F(s - )$ so close to $1 - F(s - )$ that $x(1 - F)^{-1}(y) > s$. Then for large $s$, $1 - F\{x(1 - F)^{-1}(y)\} \geq (1 - \delta)y$; whence $1 - F(s) \geq (1 - \delta)y$. Therefore, $1 - F(s) \geq (1 - \delta)\{1 - F(s - )\}$. But $1 - F(s) \leq 1 - F(s - )$ and so $1 - F(s) \sim 1 - F(s - )$ as $s \to \infty$.

Now let $t \to \infty$ through the complement of supp $F$ and put

$$t' = \inf\{u \geq t: u \in \text{supp } F\}.$$

Then $t' \in \text{supp } F$ and so by (3.3), $1 - F(xt') \sim 1 - F(t')$ as $t$ (or equivalently, $t'$) $\to \infty$. Taking $s = t'$ in the result of the last paragraph, we see that $1 - F(t') \sim 1 - F(t'- ) = 1 - F(t)$. Therefore,

$$1 - F(xt) \geq 1 - F(xt') \sim 1 - F(t') \sim 1 - F(t'-) = 1 - F(t).$$

But $1 - F(xt) \leq 1 - F(t)$ and so $1 - F(xt) \sim 1 - F(t)$ as $t \to \infty$ through the complement of supp $F$. Result (3.2) follows from this conclusion and (3.3). □

LEMMA 3.2. *If $X_{n,n-1}/X_{nn} \to_p 1$, then $X_{n,n-k}/X_{n,n-k+1} \to_p 1$ for each $k \geq 1$ and $\{1 - F(x)\}/\{1 - F(\eta x)\} \to 0$ as $x \to \infty$ for each $0 < \eta < 1$.*

PROOF. The event $X_{n,n-k}/X_{n,n-k+1} \leq \eta$, $X_{n,n-k+1} = y$ is equivalent to $n - k$ sample values less than or equal to $\eta y$, one sample value equal to $y$, and the remaining $k - 1$ sample values greater than or equal to $y$. Arguing thus we may show that for each $0 < \eta < 1$, there exist constants $0 < C_1 < C_2 < \infty$, not depending on $n$ or $\eta$, such that

$$\{P(X_{n-k}/X_{n,n-k+1} \leq \eta, y \leq X_{n,n-k+1} \leq y + dy)$$
$$\begin{cases} \leq C_2 n^k F(\eta y)^{n-k}\{1 - F(y -)\}^{k-1}\, dF(y) \\ \geq C_1 n^k F(\eta y)^{n-k}\{1 - F(y -)\}^{k-1}\, dF(y). \end{cases}$$

Integrating over $y$ we deduce that

$$(3.4)\quad P(X_{n,n-k} \le \eta X_{n,n-k+1}) \asymp \int_{-\infty}^{\infty} n^k F(\eta y)^{n-k}\{1 - F(y-)\}^{k-1}\, dF(y),$$

where $a_n \asymp b_n$ means that $a_n/b_n$ and $b_n/a_n$ are both bounded as $n \to \infty$.

Taking $k = 1$ in (3.4), we see that $P(X_{n,n-1} \le \eta X_{nn}) \to 0$ implies

$$n\int_{-\infty}^{\infty} F(\eta y)^{n-1}\, dF(y) \to 0.$$

Noting that the factor $n$ of the integral may trivially be replaced by $Cn$, for any $C > 0$, without affecting this result, we may prove that

$$\lambda\int_0^{\infty} \exp[-\lambda\{1 - F(\eta y)\}]\, dF(y) \to 0$$

as $\lambda \to \infty$ and, hence, that

$$\int_0^{\infty} e^{-t}\, d_t\big(\lambda[1 - F\{\eta^{-1}F^{-1}(1 - \lambda^{-1}t)\}]\big) \to 0$$

as $\lambda \to \infty$. Therefore, $\lambda[1 - F\{\eta^{-1}F^{-1}(1 - \lambda^{-1}t)\}] \to 0$ for each fixed $\eta > 0$, $t > 0$. In consequence, $\{1 - F(\eta^{-1}y)\}/\{1 - F(y)\} \to 0$ as $y \to \infty$, for each $\eta > 0$, the proof being similar to that used to derive (3.3). It is now a simple matter to deduce from (3.4) that

$$P(X_{n,n-k} \le \eta X_{n,n-k+1}) \to 0$$

for each $\eta > 0$ and $k \ge 1$, from which it follows that $X_{n,n-k}/X_{n,n-k+1} \to 1$ in probability. □

Proof of Theorem 2.1. Convergence in distribution to a proper limit will not occur unless either $nB_n \to_p +\infty$ or $nB_n \to_p -\infty$. We assume $nB_n \to_p +\infty$. Then

$$(3.5)\qquad n^{-1}\sum_{i=1}^{n} I(|X_i| > \varepsilon nB_n) \to_p 0$$

for each $\varepsilon > 0$. The condition that the variables $X_i^*/nB_n$ be infinitesimal in the sense of Gnedenko and Kolmogorov [10, page 95] is

$$\max_{1\le i\le n} P(|X_i^*| > \varepsilon nB_n|\mathscr{X}) \to_p 0$$

for all $\varepsilon > 0$, which reduces to (3.5).

(i) Nonnormal limit. We may deduce from [10, page 124] that if there exists a proper nondegenerate limit distribution which is not normal, then there exist nonincreasing, nonnegative functions $A_1$ and $A_2$ defined on the positive half-line, satisfying $A_1(\infty) = A_2(\infty) = 0$, not both identically zero, and such that at continuity points $x > 0$,

$$\sum_{i=1}^{n} P(X_i^* > xnB_n|\mathscr{X}) \to_p A_1(x), \qquad \sum_{i=1}^{n} P(X_i^* \le -xnB_n|\mathscr{X}) \to_p A_2(x).$$

These two conditions are, respectively, equivalent to

$$(3.6)\quad \sum_{i=1}^{n} I(X_i > xnB_n) \to_p A_1(x), \qquad \sum_{i=1}^{n} I(X_i \le -xnB_n) \to_p A_2(x).$$

The left-hand sides in (3.6) are nonnegative integers and so $A_1, A_2$ may take only nonnegative integer values.

Suppose $A_1$ is not identically zero. Then there exist $x_1 > 0$ such that $A_1(x) \ge 1$ for $x < x_1$ and $A_1(x) = 0$ for $x > x_1$. By definition of $x_1$, $X_{nn}/(nB_n) \to_p x_1$ as $n \to \infty$. This amounts only to a definition of $B_n$, requiring that $nB_n$ be asymptotic to a constant multiple of $X_{nn}$. Therefore, we may take $nB_n = X_{nn}$, noting that this implies $x_1 = 1$.

Next we consider the size of the jump of $A_1$ at $x_1$. If the jump is of two units or more, then $X_{n,n-1}/X_{nn} \to_p 1$. It then follows from Lemma 3.2 that $X_{n,n-k}/X_{n,n-k+1} \to_p 1$ for all $k \ge 1$. This means that if the jump of $A_1$ at $x_1$ exceeds one unit, then the jump is infinite. Therefore, the jump at $x_1$ must equal one unit. That is, $A_1(x_1+) = 0$, $A_1(x_1-) = 1$.

Now consider the position $x_2$ of the next jump in $A_1$: $A_1(x_2+) = 1$, $A_1(x_2-) \ge 2$. Then $X_{n,n-1}/X_{nn} \to_p x_2$, which implies $P(X_{n,n-1}/X_{nn} > x) \to 0$ whenever $x_2 < x < x_1$ $(=1)$. By Lemma 3.1, this entails $X_{n,n-1}/X_{nn} \to_p 0$, so that $x_2 = 0$. Therefore, $A_1$ satisfies $A_1(x) = 1$ for $0 < x < 1$ and $A_1(x) = 0$ for $x > 1$.

We claim that since $A_1$ does not vanish, $A_2$ must. If not, we may apply the argument above to deduce that for some $x_0 > 0$, $A_2(x) = 1$ for $0 < x < x_0$ and $A_2(x) = 0$ for $x > x_0$. Then we have

$$(3.7)\quad X_{n,n-1}/X_{nn} \to_p 0, \qquad |X_{n2}|/|X_{n1}| \to_p 0, \qquad |X_{n1}|/X_{nn} \to_p x_0.$$

Put $Z_i = |X_i|$ and let $Z_{n1} \le \cdots \le Z_{nn}$ denote the collection $Z_1, \dots, Z_n$ arranged in order of increasing magnitude. We may deduce from (3.7) that with probability tending to 1, $(Z_{nn}, Z_{n,n-1}) = (X_{nn}, |X_{n1}|)$ or $(|X_{n1}|, X_{nn})$ and $Z_{n,n-2}/Z_{n,n-1} \to_p 0$. Application of Lemma 3.1 (to a sample of $|X_i|$'s rather than $X_i$'s) now gives $Z_{n,n-1}/Z_{nn} \to_p 0$, which contradicts $|X_{n1}|/X_{nn} \to_p x_0$.

We may therefore assume that $A_1$ has the form described two paragraphs earlier, that $A_2 \equiv 0$, and that $B_n = n^{-1}X_{nn}$. Then by Lemma 3.1, $1 - F$ is slowly varying at $+\infty$ and $|X_{n1}|/X_{nn} \to_p 0$, from which it follows that

$$(3.8)\qquad P(X > x)/P(|X| > x) \to 1$$

as $x \to \infty$. Conversely, if $1 - F$ is slowly varying at $+\infty$ and (3.8) holds, then $|X_{n1}|/X_{nn} \to_p 0$, whence it follows (if we take $B_n = n^{-1}X_{nn}$) that $A_1$ has the form described earlier and $A_2 \equiv 0$. Therefore, to complete our treatment of the case of a nonnormal limit, it suffices to prove that if $1 - F$ is slowly varying at $+\infty$ and (3.8) holds, then

$$(3.9)\qquad P\{n(\bar{X}^* - \bar{X})/X_{nn} \le x \mid \mathscr{X}\} \to_p P(N - 1 \le x),$$

where $N$ is Poisson with unit mean.

Put $Z_i = X_i^2$, write $Z$ for a generic $Z_i$ and let $Z_{n1} \le \cdots \le Z_{nn}$ denote the collection $Z_1, \dots, Z_n$ arranged in order of increasing magnitude. Then $Z_{n,n-1}/Z_{nn} \to_p 0$ and so by Lemma 3.1 the function $P(Z > y)$ is slowly varying at infinity. In consequence,

$$\left(\sum_{i=1}^{n} Z_i\right)/Z_{nn} \to_p 1$$

[8] and so

$$\left(\sum_{i=1}^{n} Z_i - Z_{nn}\right)/Z_{nn} \to_p 0. \tag{3.10}$$

Denote by $\mathscr{Y} = \{Y_1, \dots, Y_n\}$ the sample obtained from $\mathscr{X}$ on replacing $X_{nn}$ by zero, and let $\mathscr{Y}^* = \{Y_1^*, \dots, Y_n^*\}$ be a resample drawn randomly, with replacement, from $\mathscr{Y}$. We may take $\mathscr{Y}^*$ to equal $\mathscr{X}^*$ when each appearance of $X_{nn}$ in the latter is replaced by zero. Put $\overline{Y} = n^{-1}\sum Y_i$, $\overline{Y}^* = n^{-1}\sum Y_i^*$. Then

$$\begin{aligned} P\left(n|\overline{Y}^* - \overline{Y}|/X_{nn} > \varepsilon|\mathscr{X}\right) &\le \varepsilon^{-2}X_{nn}^{-2}n^2E\left\{\left(\overline{Y}^* - \overline{Y}\right)^2|\mathscr{X}\right\} \\ &= \varepsilon^{-2}X_{nn}^{-2}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 \\ &\le \varepsilon^{-2}X_{nn}^{-2}\sum_{i=1}^{n} Y_i^2 \\ &= \varepsilon^{-2}Z_{nn}^{-1}\left(\sum_{i=1}^{n} Z_i - Z_{nn}\right) \to_p 0, \end{aligned} \tag{3.11}$$

by (3.10). Since $B_n = n^{-1}X_{nn}$, then

$$\left(\overline{X}^* - \overline{X}\right)/B_n = N^* - 1 + n\left(\overline{Y}^* - \overline{Y}\right)/X_{nn},$$

where $N^*$ denotes the number of times $X_{nn}$ appears in the resample $\mathscr{X}^*$. As $n \to \infty$, $N^*$ is asymptotically Poisson-distributed with unit mean. The desired result (3.9) now follows from (3.11).

(ii) Normal limit. Weeff may deduce from [10, page 128] that there exist $\mathscr{X}$-measurable sequences $A_n$, $B_n$ such that

$$P\left\{\left(\overline{X}^* - A_n\right)/B_n \le x|\mathscr{X}\right\} \to_p \Phi(x), \qquad -\infty < x < \infty,$$

if and only if there exists an $\mathscr{X}$-measurable sequence $B_n$ such that for all $\varepsilon > 0$,

$$\sum_{i=1}^{n} P\left(B_n^{-1}|X_i^*| > \varepsilon|\mathscr{X}\right) \to_p 0,$$

$$B_n^{-2}\sum_{i=1}^{n}\left(E\left\{(X_i^*)^2 I(|X_i^*| \le \varepsilon B_n)|\mathscr{X}\right\} - \left[E\left\{X_i^* I(|X_i^*| \le \varepsilon B_n)|\mathscr{X}\right\}\right]^2\right) \to_p 1.$$

These two conditions are respectively equivalent to

$$\sum_{i=1}^{n} I(|X_i| > \varepsilon B_n) \to_p 0, \tag{3.12}$$

$$B_n^{-2} \sum_{i=1}^{n} X_i^2 I(|X_i| \le \varepsilon B_n) - n^{-1} B_n^{-2} \left\{ \sum_{i=1}^{n} X_i I(|X_i| \le \varepsilon B_n) \right\}^2 \to_p 1 \tag{3.13}$$

for all $\varepsilon > 0$. In view of (3.12), (3.13) may be rewritten as

$$B_n^{-2} \sum_{i=1}^{n} X_i^2 - n^{-1} B_n^{-2} \left( \sum_{i=1}^{n} X_i \right)^2 \to_p 1,$$

that is, $nB_n^{-2}\hat{\sigma}^2 \to_p 1$. Therefore, (3.12) and (3.13) are together equivalent to $nB_n^{-2}\hat{\sigma}^2 \to_p 1$ (which simply means that $B_n$ should be chosen asymptotic to $n^{1/2}\hat{\sigma}$) and

$$\sum_{i=1}^{n} I(X_i^2 > \varepsilon n \hat{\sigma}^2) \to_p 0$$

for all $\varepsilon > 0$. The latter relation is equivalent to

$$\Big( \max_{1 \le i \le n} X_i^2 \Big) \Big/ \sum_{i=1}^{n} (X_i - \bar{X})^2 \to_p 0. \tag{3.14}$$

Put $U_n = (\max_{i \le n} X_i^2)/(\sum_{i \le n} X_i^2)$. Since $\sum (X_i - \bar{X})^2 \le \sum X_i^2$, then (3.14) implies $U_n \to_p 0$. The converse is clear if $E(X^2) < \infty$. We show next that when $E(X^2) = \infty$, $U_n \to_p 0$ implies (3.14). For this it suffices to prove that for some $\varepsilon > 0$,

$$\left\{ \sum_{i=1}^{n} (X_i - \bar{X})^2 \right\} \Big/ \left( \sum_{i=1}^{n} X_i^2 \right) \ge \varepsilon + o_p(1) \tag{3.15}$$

as $n \to \infty$. Choose $\lambda > 0$ so large that $P(|X| > \lambda) < 1$ and put

$$S_n = n^{-1} \sum_{i=1}^{n} X_i^2 I(|X_i| > \lambda), \qquad T_n = n^{-1} \sum_{i=1}^{n} I(|X_i| > \lambda).$$

Then

$$\bar{X}^2 \le \left( n^{-1} \sum_{i=1}^{n} |X_i| \right)^2 \le \left\{ \lambda + n^{-1} \sum_{i=1}^{n} |X_i| I(|X_i| > \lambda) \right\}^2 \le \left\{ \lambda + (S_n T_n)^{1/2} \right\}^2.$$

Therefore,

$$\begin{aligned} n^{-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 &= n^{-1} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 \\ &\ge S_n - \left\{ \lambda^2 + 2\lambda (S_n T_n)^{1/2} + S_n T_n \right\} \\ &= S_n \left\{ 1 - \lambda^2 S_n^{-1} - 2\lambda (T_n / S_n)^{1/2} - T_n \right\} \\ &= S_n \left\{ 1 - P(|X| > \lambda) + o_p(1) \right\}, \end{aligned} \tag{3.16}$$

since $S_n \to_p \infty$ and $T_n \to_p P(|X| > \lambda)$. Also,

$$S_n = n^{-1} \sum_{i=1}^{n} X_i^2 - E\{X^2 I(|X| \le \lambda)\} + o_p(1).$$

Result (3.15) now follows from (3.16).

At this point we need a portion of Lemma 3.3, due to O'Brien [13].

LEMMA 3.3. *Let $Y, Y_1, Y_2, \ldots$ be independent and identically distributed random variables with $P(Y \ge 0) = 1$ and $P(Y > 0) > 0$. Put $U_n = (\max_{i \le n} Y_i)/(\sum_{i \le n} Y_i)$. Then $U_n \to_p 0$ if and only if $E\{YI(Y \le y)\}$ is slowly varying at $\infty$ and $U_n \to 0$ almost surely if and only if $E(Y) < \infty$.*

Note that $X$ is in the domain of attraction of the normal law if and only if $E\{X^2 I(|X| \le x)\}$ is slowly varying [12, page 84]. This condition is in turn equivalent to the slow variation of $E\{X^2 I(X^2 \le x)\}$ and hence (by Lemma 3.3) to $U_n \to_p 0$. We have already shown that the asymptotic normality of the bootstrap mean is equivalent to $U_n \to_p 0$, and so also to $X$ being in the domain of attraction of the normal law.

We have already shown that $B_n^2 = \hat{\sigma}^2$ suffices as the scaling constant. A suitable version of $A_n$ is

$$A_n = n^{-1} \sum_{i=1}^{n} E\{X_i^* I(X_i^* \le B_n) | \mathscr{X}\} = n^{-1} \sum_{i=1}^{n} X_i I(|X_i| \le B_n).$$

In view of (3.12), this implies $A_n = \bar{X}$ will also do. □

PROOF OF THEOREM 2.2. Let $Z_n$ be a nonnegative function of the sample $\mathscr{X}$. Denote by $\mathscr{Y} = \{Y_1, \ldots, Y_n\}$ the sample obtained from $\mathscr{X}$ on replacing $X_i$ by zero whenever $|X_i| > Z_n$, and otherwise leaving $X_i$ unchanged for $1 \le i \le n$. Put $\bar{Y} = n^{-1} \sum Y_i$ and $\bar{Y}^* = n^{-1} \sum Y_i^*$, where $\{Y_1^*, \ldots, Y_n^*\}$ is a resample drawn randomly, with replacement, from $\mathscr{Y}$. We claim that for each $\varepsilon > 0$ there exists $\delta > 0$, depending only on $\varepsilon$, on $\limsup nP(|X| > c_n)$ and on the distribution of $X$, such that if $P(c_n^{-1} Z_n > \delta) < \delta$ for all large $n$, then

$$\limsup_{n \to \infty} P\{P(nc_n^{-1}|\bar{Y}^* - \bar{Y}| > \varepsilon | \mathscr{X}) > \varepsilon\} \le \varepsilon. \tag{3.17}$$

To verify this claim, observe that

$$n^2 E\{(\bar{Y}^* - \bar{Y})^2 | \mathscr{X}\} \le nE\{(Y_1^*)^2 | \mathscr{X}\} = \sum_{i=1}^{n} X_i^2 I(|X_i| \le Z_n).$$

Therefore by Markov's inequality, for each $u > 0$,

$$\begin{aligned} &P(nc_n^{-1}|\bar{Y}^* - \bar{Y}| > \varepsilon | \mathscr{X}) \\ &\quad \le I(c_n^{-1} Z_n > u) + I(c_n^{-1} Z_n \le u) \varepsilon^{-2} c_n^{-2} E\{(\bar{Y}^* - \bar{Y})^2 | \mathscr{X}\} \\ &\quad \le I(c_n^{-1} Z_n > u) + \varepsilon^{-2} c_n^{-2} \sum_{i=1}^{n} X_i^2 I(|X_i| \le c_n u). \end{aligned}$$

In consequence,

$$P\{P(nc_n^{-1}|\bar{Y}^* - \bar{Y}| > \varepsilon|\mathscr{X}) > \varepsilon\}$$

$$\leq P(c_n^{-1}Z_n > u) + P\left\{\varepsilon^{-2}c_n^{-2}\sum_{i=1}^{n} X_i^2 I(|X_i| \leq c_n u) > \varepsilon\right\} \tag{3.18}$$

$$\leq P(c_n^{-1}Z_n > u) + \varepsilon^{-3}c_n^{-2}nE\{X^2 I(|X| \leq c_n u)\}.$$

Choose $\delta \leq (1/2)\varepsilon$ so small that

$$C_1\varepsilon^{-3}\left\{\limsup_{n\to\infty} nP(|X| > c_n)\right\}\limsup_{x\to\infty}\{\delta^2 P(|X| > x\delta)/P(|X| > x)\} \leq (1/2)\varepsilon$$

and take $u = \delta$ in (3.18). By hypothesis, the right-hand side of (3.18) is dominated by

$$\delta + \varepsilon^{-3}c_n^{-2}nC_1(c_n\delta)^2 P(|X| > c_n\delta)$$

$$\leq \delta + C_1\varepsilon^{-3}\{nP(|X| > c_n)\}\{\delta^2 P(|X| > c_n\delta)/P(|X| > c_n)\}$$

$$\leq \delta + (1/2)\varepsilon + o(1) \leq \varepsilon + o(1).$$

This proves (3.17).

Take $Z_n = Z_n(r)$ to be $|X_{\langle n, r+1\rangle}|$ and let $N$ denote the number of $|X_i|$'s, $1 \leq i \leq n$, which exceed $\delta c_n$. Then

$$P(c_n^{-1}Z_n > \delta) = P(N \geq r) \leq r^{-1}E(N) = r^{-1}nP(|X| > \delta c_n),$$

from which it follows that for each $\delta > 0$,

$$\lim_{r\to\infty}\limsup_{n\to\infty} P\{c_n^{-1}Z_n(r) > \delta\} = 0. \tag{3.19}$$

Note too that for our choice of $Z_n$,

$$\sum_{i=1}^{n} X_i^* = \sum_{i=1}^{r} N_i^* X_{\langle n,i\rangle} + \sum_{i=1}^{n} Y_i^*,$$

whence

$$nc_n^{-1}(\bar{X}^* - \bar{X}) = nc_n^{-1}\sum_{i=1}^{r}(N_i^* - 1)X_{\langle n,i\rangle} + nc_n^{-1}(\bar{Y}^* - \bar{Y}).$$

Therefore, $\Delta_{n,r} = nc_n^{-1}(\bar{Y}^* - \bar{Y})$. The theorem now follows from (3.17) and (3.19). □

**Acknowledgments.** I am grateful to Michael J. Klass for helpful conversations. The referee's detailed comments have also been most helpful.

## REFERENCES

[1] Arov, D. Z. and Bobrov, A. A. (1960). The extreme terms of a sample and their role in the sum of independent variables. *Theory Probab. Appl.* **5** 377–396.

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

[3] ATHREYA, K. B. (1987). Bootstrap of the mean in the infinite variance case. II. Unpublished.

[4] BINGHAM, N. H. and TEUGELS, J. L. (1981). Conditions implying domains of attraction. In *Proc. Sixth Conf. Probab. Theory* (B. Bereanu, Ş. Grigorescu, M. Iosifescu and T. Postelnicu, eds.) 23–34. Editura Academiei, Bucharest.

[5] CSÖRGŐ, S. (1987). Notes on extreme and self-normalized sums from the domain of attraction of a stable law. Preprint No. 83/1987, Math. Inst., Hungar. Acad. Sci.

[6] CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986). Weighted empirical and quantile processes. *Ann. Probab.* **14** 31–85.

[7] CSÖRGŐ, M., CSÖRGŐ, S., HORVÁTH, L. and MASON, D. M. (1986). Normal and stable convergence of integral functions of the empirical distribution function. *Ann. Probab.* **14** 86–118.

[8] DARLING, D. A. (1952). The influence of the maximum term in the addition of independent random variables. *Trans. Amer. Math. Soc.* **73** 95–107.

[9] GALAMBOS, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed. Krieger, Malabar, Fla.

[10] GNEDENKO, B. V. and KOLMOGOROV, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley, Reading, Mass.

[11] HALL, P. (1978). On the extreme terms of a sample from the domain of attraction of a stable law. *J. London Math. Soc.* (2) **18** 181–191.

[12] IBRAGIMOV, I. A. and LINNIK, YU. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

[13] O'BRIEN, G. L. (1980). A limit theorem for sample maxima and heavy branches in Galton–Watson trees. *J. Appl. Probab.* **17** 539–545.

[14] RESNICK, S. I. and RUBINOVITCH, M. (1973). The structure of extremal processes. *Adv. in Appl. Probab.* **5** 287–307.

[15] SMID, B. and STAM, A. J. (1975). Convergence in distribution of quotients of order statistics. *Stochastic Process. Appl.* **3** 287–292.

[16] SMIRNOV, N. V. (1952). Limit distributions for terms of a variational series. Amer. Math. Soc. Translation No. 67.

STATISTICS RESEARCH SECTION
SCHOOL OF MATHEMATICAL SCIENCES
MATHEMATICAL SCIENCES BUILDING
AUSTRALIAN NATIONAL UNIVERSITY
G.P.O. BOX 4
CANBERRA, A.C.T. 2601
AUSTRALIA