Statistica Sinica 8(1998), 887-906

COMPARISON OF BOOTSTRAP AND ASYMPTOTIC APPROXIMATIONS TO THE DISTRIBUTION OF A HEAVY-TAILED MEAN

Peter Hall and Bing-Yi Jing

Australian National University

Abstract: It is well-know that for heavy-tailed distributions the bootstrap can lead to inconsistent estimation of the distribution of the sample mean; and that this difficulty may be overcome by using the so-called "subsample bootstrap", where the size of a bootstrap resample is an order of magnitude smaller than that of the sample. Naturally, one might ask whether, as in classical problems, the bootstrap applied to heavy-tailed distributions produces more accurate approximations to the distribution of the sample mean than do asymptotic methods. We show that, generally speaking, it does not. In an important class of problems, the subsample bootstrap performs more poorly than asymptotic methods, even if the subsample size is chosen optimally. A technique related to Richardson extrapolation, effectively a cross between the subsample bootstrap and asymptotic methods, performs better than either approach in some, but not all, circumstances.

Key words and phrases: Asymptotic approximation, bootstrap, central limit theorem, convergence rate, domain of attraction, Edgeworth expansion, mean, percentile, pivot, Richardson extrapolation, stable law.

1. Introduction

When a sampling distribution has heavy tails, the usual form of the bootstrap fails to consistently estimate the distribution of the sample mean. The reason is that in the case of heavy-tailed distributions, the size of the sample mean is determined by the values of a small number of extreme order statistics; and the usual bootstrap does not consistently estimate the distributions of extreme values. Athreya (1987) showed that this problem may be alleviated by employing the "subsample bootstrap", where resamples have a size which is an order of magnitude smaller than that of the original sample.

This result does not, however, indicate whether the "subsample bootstrap" provides a more accurate estimate of the true distribution of the mean than do more standard, asymptotic methods. In the present paper we address this issue, by developing second-order theory describing the subsample bootstrap in the context of heavy-tailed distributions. We show that the bootstrap is not second-order accurate, in that it fails to correct for the first term describing departure from the limit distribution. Worse than this, even when the subsample size is chosen optimally, the error between the subsample bootstrap approximation and the true distribution is often an order of magnitude larger than that of an asymptotic approximation. Therefore, while neither the bootstrap nor the asymptotic approximation succeeds in capturing the first term in an Edgeworth-type expansion of error, the asymptotic approach is considerably more accurate.

These results are admittedly for a percentile, non-pivotal version of the bootstrap. However, in the case of heavy-tailed distributions where the tail exponent is unknown, a pivotal approximation to the distribution of the sample mean cannot be achieved simply by correcting for scale, and so pivoting is not nearly as attractive as it is in more conventional settings. We do suggest a two-parameter approach to pivoting, but point out that while it does provide one-term Edgeworth correction it is computationally expensive, and its properties depend intimately on the context to which it is applied.

On the other hand, we show that a hybrid approach, based quite literally on a mixture of asymptotic and subsample bootstrap methods, can produce improved performance. This method is applicable in a wide variety of settings, where the subsample bootstrap is used in order to overcome problems of inconsistency of standard bootstrap methods. (In this context, see Huang, Sen and Shao (1996).) The main requirement is that the second-order correction supplied by the subsample bootstrap be accurate except for a known or estimable factor $c(m,n)$, depending on both subsample size m and sample size n. (If $c(m,n)$ represents the amount by which the second-order term in the subsample bootstrap should be multiplied in order to equal the true second-order term, then it generally converges to zero as n increases.) The proportions in which the two approximations should be mixed depend only on $c(m,n)$. Applications include, for example, estimation of the distribution of an estimate of an extremal eigenvalue of a matrix, when one or more eigenvalues are tied for that value. Bickel and Yahav (1988) were apparently the first to use the method in a bootstrap setting. The "m out of n" bootstrap has been studied by, for example, Beran and Srivastava (1985), Hall (1990a), Wu (1990), Mammen (1992), Politis and Romano (1992, 1994), Politis, Romano and You (1993), Bickel, Götze and van Zwet (1994) and authors cited therein, usually in the context of applications where the classical form of the bootstrap does not produce consistency.

In addition to the pioneering work of Athreya (1987), first-order properties of the subsample bootstrap for heavy-tailed distributions have been studied by Giné and Zinn (1989), Knight (1989) and Hall (1990b). Arcones and Giné (1989, 1992) have investigated related issues, including consistency of both distribution and moment estimators based on the bootstrap, for sums from quite general distributions; and Deheuvels, Mason and Shorack (1993) have addressed the influence of extremes on the bootstrap. Asymptotic expansions of the distribution of a mean when the sampling distribution lies in the domain of normal attraction of a stable law have been developed by Cramér (1962, 1963), Zolotarev (1962) and Hall (1981). Our main result on bootstrap methods, Theorem 2.2, applies to a subset of each of the contexts described by these authors.

Section 2 will describe our main theoretical results, and Section 3 will summarize a simulation study which, for moderate sample sizes, illustrates our theoretical conclusions. Proofs of results in Section 2 are outlined in Section 4.

2. Main Results

2.1. Introduction and summary

Let $Y, Y_1, Y_2, \ldots$ denote independent and indentically distributed random variables in the domain of normal attraction of a symmetric stable law $H = H(\cdot|\alpha, a)$, with exponent α and scale parameter a, where $0 < \alpha < 2$, $\alpha \neq 1$ and $a > 0$. Then H has characteristic function $\exp(-a|t|^{\alpha})$, and a may be chosen so that the distribution of $S_n = n^{-1/\alpha} \sum_{j \leq n}(Y_j - \mu)$ converges to H, where $\mu = 0$ if $0 < \alpha < 1$ and $\mu = E(Y)$ if $1 < \alpha < 2$. We wish to estimate the distribution function G_n of S_n.

The asymptotic approach to this problem is to estimate H, and so the performance of the asymptotic estimator depends at least in part on the rate of convergence of G_n to H. This is perhaps best expressed in terms of a short Edgeworth expansion, $G_n = H_n + o(\delta_n)$ as $n \to \infty$, where $\{\delta_n\}$ denotes a sequence of positive constants decreasing to zero and H_n is an Edgeworth expansion of which the first term is H. The rate of convergence of the asymptotic approximation will not be faster than the order of $H_n - H$. Expansions have been developed by Cramér (1962, 1963) and Zolotarev (1962), for example.

Bootstrap methods offer an alternative way of approximating G_n, and take different forms in the cases $\alpha < 1$ and $\alpha > 1$. To describe them, let $\mathcal{X} = \{Y_1, \ldots, Y_n\}$ denote the first n of the variables Y_j introduced two paragraphs earlier, and let $Y_1^*, Y_2^*, \ldots$ be random variables obtained by sampling randomly with replacement from $\mathcal{X}$. Define $Z_i^* = Y_i^*$ when $\alpha < 1$, and $Z_i^* = Y_i^* - \bar{Y}$ when $\alpha > 1$, where $\bar{Y} = n^{-1} \sum_{j \leq n} Y_j$. Put $S_m^* = m^{-1/\alpha} \sum_{j \leq m} Z_j^*$ and write $\widehat{G}_m$ for the distribution function of S_m^* conditional on $\mathcal{X}$. Then $\widehat{G}_m$ is the percentile bootstrap estimator of G_m. It shares the limit, H, of G_n if and only if $m = m(n)$ diverges in such a way that $m/n \to 0$ as $n \to \infty$.

Subsection 2.2 describes the relative performance of asymptotic and bootstrap approximations, pointing out that the former is usually preferable. This is the central conclusion of our paper. A hybrid approach which can, on occasion, perform better than either asymptotic or Edgeworth methods, is also discussed in subsection 2.2.

The theoretical results that underpin these conclusions are introduced in subsections 2.3 and 2.4. Subsection 2.3 develops a typical Edgeworth expansion H_n. For the sake of brevity, only the case $\alpha > 1$ is be treated there. (Results and proofs for $\alpha < 1$ are very similar, but that case does not feature in our comparison of bootstrap and asymptotic methods in subsection 2.3.) Approximation theory for $\widehat{G}_m$ is developed in subsection 2.4. There we treat both the cases $\alpha < 1$ and $\alpha > 1$, since there are significant technical differences between them.

2.2. Practical implications

For the sake of definiteness we assume that the sampling distribution arises as an inverse power of a continuous distribution. That is, $Y_1, \ldots, Y_n$ are distributed as $Y = (\mathrm{sgn}\, X)|X|^{-1/\alpha}, 1 < \alpha < 2$, where X has a continuous distribution. The practical problem which we shall address is that of estimating the distribution of $\bar{Y} - \mu$; in effect, of estimating G_n. We show that, (a) even for optimal choice of m, the subsample bootstrap is outperformed by a relatively simple asymptotic approximation in which Hill's (1975) method is employed to estimate the unknowns α and a; and (b) the asymptotic approximation approach can, on occasion, be improved by combining it with the bootstrap.

We begin by describing the asymptotic method, which we claim produces an approximation to G_n that is in error by terms of order $\xi_n = n^{1-(2/\alpha)} \vee (n^{-2/5} \log n)$. Provided that the distribution function F of X has three continuous derivatives in a neighbourhood of the origin, $P(|Y| > y) = 2y^{-\alpha} F'(0) + \frac{1}{3} y^{-3\alpha} F'''(0) + o(y^{-3\alpha})$ as $y \to \infty$. In this circumstance, a technique based on extreme value theory (Hill (1975)) may be used to develop very simple estimators of α and $F'(0)$, with convergence rates $n^{-2/5}$ and $n^{-2/5} \log n$ respectively (Hall (1982), Csörgő, Deheuvels and Mason (1985)). Faster rates may be achieved using more sophisticated approaches, but generally, a convergence rate of $n^{-1/2}$ cannot be achieved without parametric knowledge.

The distribution of $S_n = n^{-1/\alpha} \sum_{j \le n} (Y_j - \mu)$ converges to $H(\cdot|\alpha, a)$, with characteristic function $\exp(-a|t|^\alpha)$, where

$$a = a_1 = 2F'(0) \int_0^\infty x^{-\alpha} \sin x dx = 2F'(0)\pi / \{\Gamma(\alpha) \sin(\alpha\pi/2)\} > 0. \qquad (2.1)$$

Substituting the estimators just above into this formula we obtain an estimator of a that converges at rate $n^{-2/5} \log n$. Therefore, the estimator $\widehat{H} = H(\cdot|\hat{\alpha}, \hat{a})$ of $H = H(\cdot|\alpha, a)$ is in error by terms of size $n^{-2/5} \log n$. In Theorem 2.1 we shall point out that $G_n - H$ is of size $n^{1-(2/\alpha)}$. It follows that if we take $\widehat{H}$ to be our approximation to G_n then the error is of size ξ_n. This result continues to hold if the norming exponent $1/\alpha$, in $n^{1/\alpha}$, is estimated as well — substituting $\hat{\alpha}$ for α there produces another error term of order ξ_n.

Next we note that in the same setting, and for optimal choice of m, the bootstrap suffers an error of size $\eta_n = n^{-(\alpha-1)(2-\alpha)/\alpha}$, which is of larger order than ξ_n. In subsection 2.4 we shall argue that the bootstrap fails to correct for the first term in an Edgeworth expansion of G_n, and in fact adds its own error from the same source, of size $m^{1-(2/\alpha)}$. There is additionally an error of $(mn^{-1})^{1-(1/\alpha)}$ arising from the bootstrap approximation of G_m by $\widehat{G}_m$; see (2.15). The over-all error is minimized by choosing m so that these two quantities are of the same order, which entails $m \simeq n^{\alpha-1}$ and gives an over-all error of size η_n.

There is an alternative bootstrap method, which mixes both the techniques discussed above. It is based on "correcting" $\widehat{G}_m$ so as to make the first term in its Edgeworth expansion coincide with that of G_n. Specifically, let $\hat{\alpha}$ and $\hat{a}$ denote the Hill estimators of α and a, write $\widehat{H} = H(\cdot|\hat{\alpha}, \hat{a})$ for the asymptotic approximation, and put

$$\widetilde{G}_n = \{1 - (mn^{-1})^{(2/\hat{\alpha})-1}\}\widehat{H} + (mn^{-1})^{(2/\hat{\alpha})-1}\widehat{G}_m. \tag{2.2}$$

This is motivated by an Edgeworth expansion, $G_n(x) = H(x) + n^{1-(2/\alpha)}A_1(x) +$ smaller order terms, which we shall give in Theorem 2.1. In view of this expansion, if H and G_m (rather than G_n) were known, a "Richardson extrapolation" argument would suggest

$$\{1 - (mn^{-1})^{(2/\alpha)-1}\}H + (mn^{-1})^{(2/\alpha)-1}G_m \tag{2.3}$$

as an approximation to G_n. Replacing α, H and G_m in this formula by their estimators we obtain the estimator at (2.2).

The dominant terms in $\widetilde{G}_n - G_n$ are of sizes $n^{-2/5}\log n$, arising from the approximation of H by $\widehat{H}$, and $(mn^{-1})^{(2/\alpha)-1}\{(mn^{-1})^{1-(1/\alpha)} + (m^{-1/\alpha} \vee m^{2-(4/\alpha)})\}$, coming from the approximation of G_m by $\widehat{G}_m$. This suggests taking $m \approx n^{1-(1/\alpha)}$ for $\alpha \le 3/2$ and $m \approx n^{(\alpha-1)/(3-\alpha)}$ for $\alpha > 3/2$, which produces an over-all error rate of $n^{-2/5}\log n$ when $\alpha \le 4-\sqrt{6} = 1.55$ and $n^{-2(2-\alpha)/\alpha(3-\alpha)}$ when $\alpha > 4-\sqrt{6}$. In the range $10/7 < \alpha < 2$ this improves on the performance of the straight asymptotic approximation method, discussed earlier.

This hybrid or mixture approach is new in the context of the bootstrap for heavy-tailed means, but has been used before in other settings. In its general form it involves approximating an unknown distribution, Ψ say, by two others that are known, Ψ_0 and Ψ_1. If

$$\Psi = \Psi_0 + \epsilon\psi + o(\epsilon) \text{ and } \Psi_1 = \Psi_0 + \delta\psi + o(\delta)$$

for known constants ϵ and δ, and an unknown function ψ, then

$$\{1 - (\epsilon/\delta)\}\Psi_0 + (\epsilon/\delta)\Psi_1 \tag{2.4}$$

is often a better approximation to Ψ_1 than was Ψ_0. In our work the roles of $\Psi, \Psi_1, \Psi_0, \epsilon$, and δ are played by $G_n, G_m, H, n^{1-(2/\alpha)}$ and $m^{1-(2/\alpha)}$, respectively; and (2.4) is equivalent to (2.3). Further examples include the case where Ψ is the distribution of a sum of independent random variables with finite, nonzero third moment, and Ψ_0, Ψ_1 are standard normal and chi-squared distribution functions, respectively. There, ϵ/δ is proportional to the ratio of two different skewnesses (with that for Ψ estimated from data, if its true value is not known). In an application (to the Studentized bootstrap) of the efficient simulation method proposed by Bickel and Yahav (1988), Ψ_0 would be as before, and Ψ_1 would be a Monte Carlo distribution function approximation obtained using bootstrap samples of smaller size than the original sample. In this case the smaller resample size is not necessarily employed because the bootstrap fails when the full sample size is used, but because resampling with a smaller resample is less computationally expensive. However, any application of the bootstrap where smaller resamples are necessary for consistency, such as the context studied in this paper, is a potential problem to which to apply the hybrid method. Beyond the bootstrap, Booth and Hall (1993) have employed the technique in the context of jackknife estimation of a distribution function.

If both the parameters α and a are unknown then there is no simple analogue of the percentile-t bootstrap, which has found favour in more classical problems concerning the bootstrap (e.g. Hall (1988)). One approach to pivoting would be as follows. Let $\hat{\alpha}, \hat{a}$ denote estimators of α, a, and consider boostrapping $T = H\{n^{1-(1/\hat{\alpha})}(\bar{Y} - \mu)|\hat{\alpha}, \hat{a}\}$ instead of $\bar{Y} - \mu$ or $\bar{Y}$. Since the asymptotic distribution of T does not depend on unknowns — it is the uniform distribution on the interval $(0, 1)$ — then this approach will correct for the first term in an Edgeworth expansion. However, it is time consuming to implement.

2.3. Properties of the unconditional distribution

In the case $1 < \alpha < 2$ we use a simple motivating example, involving a distribution expressible as a negative power of a continuous random variable X (compare Hall (1981)). More general Edgeworth expansions will be discussd in Remark 2.3.

Put $Y = (\operatorname{sgn} X)|X|^{-1/\alpha}$, where $1 < \alpha < 2$, and let $\mu = E(Y)$ and $Z = Y - \mu$. Assume that

> for $1 < \alpha \le 4/3$ the distribution function F of X has two derivatives in a neighbourhood of the origin, and F'' satisfies a Lipschitz condition of order $\epsilon > 0$; that for $4/3 < \alpha < 2$, F has one derivative in a neighbourhood of the origin, and F' satisfies a Lipschitz condition of order $(4/\alpha) - 2 + \epsilon$ for some $\epsilon > 0$; and that $F'(0) > 0$. (2.5)

This condition is satisified if F has three bounded derivatives in a neighbourhood of the origin, and $F'(0) > 0$. Define $a = a_1$ as in (2.1),

$$a_2' = \int_0^\infty x\{P(|Y| > x) - 2F'(0)x^{-\alpha}\}dx,$$
$$a_3 = \begin{cases} F''(0)\int_0^\infty x^{-2\alpha}(1-\cos x)dx & \text{if } 1 < \alpha \le 4/3 \\ 0 & \text{if } 4/3 < \alpha < 2, \end{cases}$$

$a_2 = a_2' - \frac{1}{2}\mu^2$. Let $L = A_j$, B_1, B_2 denote the real-valued functions whose respective Fourier-Stieltjes transforms, $\int_{-\infty}^\infty e^{itx}dL(x)$, are given by $(1/j!)(-a_2t^2)^j$ $\exp(-a_1|t|^\alpha)$, $ia_1\mu(\operatorname{sgn} t)|t|^{\alpha+1}\exp(-a_1|t|^\alpha)$, $-(\frac{1}{2}a_1^2+ia_3\operatorname{sgn} t)|t|^{2\alpha}\exp(-a_1|t|^\alpha)$. In this notation, $A_0 = H$, the distribution function of the limiting stable law. Write G_n for the distribution function of $S_n = n^{-1/\alpha}\sum_{j\le n} Z_j$.

Theorem 2.1. *Under condition* (2.5),

$$G_n(x) = \sum_{j=0}^{2} n^{j\{1-(2/\alpha)\}}A_j(x)+n^{-1/\alpha}B_1(x)+n^{-1}B_2(x)+o(n^{-1}+n^{2-(4/\alpha)}) \quad (2.6)$$

uniformly in x as $n \to \infty$.

Remark 2.1. Rate of convergence

The rate of convergence in this limit theorem can be made slower than $n^{-\epsilon}$, for any given $\epsilon > 0$, by choosing α sufficiently close to 2; note the presence of the term in $n^{1-(2/\alpha)}$ in the expansion

$$H_n(x) = \sum_{j=0}^{2} n^{j\{1-(2/\alpha)\}}A_j(x) + n^{-1/\alpha}B_1(x) + n^{-1}B_2(x),$$

representing the non-remainder portion of the right-hand side of (2.6). If follows that an asymptotic approximation to the distribution function G_n, such as that suggested in subsection 2.1, may be in error by more than order $n^{-\epsilon}$ regardless of the accuracy with which we may estimate the unknowns α and a.

Remark 2.2. Short Edgeworth expansions

If $\mu \ne 0$ then a simplified three-term version of formula (2.6) may be expressed as

$$G_n(x) = \begin{cases} A_0(x) + n^{1-(2/\alpha)}A_1(x) \\ \quad +n^{-1/\alpha}B_1(x) + o(n^{-1/\alpha}) & \text{if } 1 < \alpha < 3/2 \\ A_0(x) + n^{-1/3}A_1(x) \\ \quad +n^{-2/3}\{A_2(x) + B_1(x)\} + o(n^{-2/3}) & \text{if } \alpha = 3/2 \\ A_0(x) + n^{1-(2/\alpha)}A_1(x) \\ \quad +n^{2-(4/\alpha)}A_2(x) + o(n^{2-(4/\alpha)}) & \text{if } 3/2 < \alpha < 2. \end{cases}$$

However, shoule μ vanish then so does B_1, and these formulae should be replaced by

$$G_n(x) = \begin{cases} A_0(x) + n^{1-(2/\alpha)}A_1(x) \\ \quad +n^{-1}B_2(x) + o(n^{-1}) & \text{if } 1 < \alpha < 4/3 \\ A_0(x) + n^{1-(2/\alpha)}A_1(x) \\ \quad +n^{-1}\{A_2(x) + B_2(x)\} + o(n^{-1}) & \text{if } \alpha = 4/3 \\ A_0(x) + n^{1-(2/\alpha)}A_1(x) \\ \quad +n^{2-(4/\alpha)}A_2(x) + o(n^{2-(4/\alpha)}) & \text{if } 4/3 < \alpha < 2. \end{cases}$$

Remark 2.3. More general Edgeworth expansions

Expansions in general contexts are easily derived using arguments similar to those employed to establish Theorem 2.1. For example, suppose independent and identically distributed random variables Z_j have common characteristic function ψ; let π_n be a complex-valued function such that the inverse Fourier-Stieltjes transform H_n of $\exp(-a|t|^\alpha)\pi_n(t)$ is real-valued; and assume that

$$\lim_{t\to 0} |t|^{-\alpha}\{1-\psi(t)\} = a > 0, \quad \sup_{|t|>\epsilon} |\psi(t)| < 1, \tag{2.7}$$

$$\pi_n(0) = 1, \qquad \sup_{-\infty<t<\infty} (1+|t|)^{-C_1}|\pi_n(t)| \le C_2 < \infty, \tag{2.8}$$

$$\int_{|t|\le n^\xi} |t|^{-1}|\psi(t/n^{1/\alpha})^n - \exp(-a|t|^\alpha)\pi_n(t)|dt = o(\delta_n) \tag{2.9}$$

as $n \to \infty$, for all $\epsilon > 0$ and some $\xi > 0$, where a, C_1, C_2, δ_n are positive and $\delta_n \to 0$. (A candidate for H_n— and hence π_n — is noted in Remark 2.1, but of course there are other possibiliteis.) Then for all $K > 0$, the distribution function G_n of $n^{-1/\alpha}\sum_{j\le n} Z_j$ satisfies

$$G_n(x) = P(S_n \le x) = H_n(x) + o(\delta_n) + O(n^{-K}) \tag{2.10}$$

uniformly in x. Condition (2.5) serves to ensure that (2.7)-(2.9) hold with $a = a_1$, $\delta_n = n^{-1} + n^{2-(4/\alpha)}$ and π_n given by (4.5) (see Section 4). In the next subsection we shall assume a very slightly stronger version of (2.9): for some ξ, $C > 0$,

$$\int_{|t|\le n^\xi} |t|^{-1}|\psi(t/n^{1/\alpha})^n - \exp(-a|t|^\alpha)\pi_n(t)|^2\exp(C|t|^\alpha)dt = o(\delta_n^2), \tag{2.11}$$

which again follows form (2.5).

2.4. Properties of the conditional distribution

In this subsection we shall develop theory describing the bootstrap distribution function $\widehat{G}_m$, introduced in subsection 2.1. Our context will be the general one discussed in Remark 2.3. Thus, we write H_n for a real-valued function (an Edgeworth expansion of G_n) whose Fourier-Stieltjes transform is $\exp(-a|t|^\alpha)\pi_n(t)$.

Define $\bar{Z} = n^{-1}\sum_{j\le n} Z_j$, $\widetilde{H}_m = H_m$ if $0 < \alpha < 1$, and $\widetilde{H}_m(x) = H_m(x + m^{1-(1/\alpha)}\bar{Z})$ if $1 < \alpha < 2$. In both cases put

$$M_m(x) = mn^{-1}\sum_{j=1}^{n}\{H_m(x - m^{-1/\alpha}Z_j) - EH_m(x - m^{-1/\alpha}Z_j)\}.$$

As we shall show in Remarks 2.4-2.6, $\widetilde{H}_m - H_m$ and M_m together represent the dominant contributions to the error, $\widehat{G}_m - G_m$, of the bootstrap approximation.

Let $\{c_n\}$ denote any increasing sequence of constants such that $\sum_{n\ge 2} (nc_n \log n)^{-1} < \infty$. For example, we might take $c_n = (\log n)^\epsilon$ for arbitrary $\epsilon > 0$.

Theorem 2.2. *Assume* (2.7), (2.8) *and* (2.11) *hold, and that* $mn^{-1}(\log n)^2 + m^{-\epsilon}\log n \to 0$ *for all* $\epsilon > 0$. *Then*

$$\widehat{G}_m(x) = \widetilde{H}_m(x) + M_m(x) + o(\delta_m) + \begin{cases} O[mn^{-1}\log n + \{(mn)^{-1}\log n\}^{1/2}] & if\ 0<\alpha<1 \\ O[(mn^{-1})^{2-(2/\alpha)}(c_n \log n)^{2/\alpha} \\ \quad +\{(mn)^{-1}\log n\}^{1/2}] & if\ 1<\alpha<2 \end{cases} \tag{2.12}$$

with probability one, and

$$\hat{G}_m(x) = \tilde{H}_m(x) + M_m(x) + o_p(\delta_m) + \begin{cases} O_p\{mn^{-1} + (mn)^{-1/2}\} & if\ 0<\alpha<1 \\ O_p\{(mn^{-1})^{2-(2/\alpha)} + (mn)^{-1/2}\} & if\ 1<\alpha<2. \end{cases} \tag{2.13}$$

We shall prove only (2.12), in Section 3, since a derivation of (2.13) is similar but simpler.

By (2.10) and either (2.12) or (2.13),

$$\widehat{G}_m - G_m = \widetilde{H}_m - H_m + M_m + \text{remainder terms.} \tag{2.14}$$

In the remarks below we show that the remainder terms in (2.14) are genuinely negligible relative to $\widetilde{H}_m - H_m$ and M_m. The remarks will also derive the orders of $\widetilde{H}_m - H_m$ and M_m, showing that

$$\widehat{G}_m(x) - G_m(x) = \begin{cases} (mn^{-1})^{1-(1/\alpha)}H'(x)S_n \\ \quad +o_p\{\delta_m + (mn^{-1})^{1-(1/\alpha)}\} & \text{if } 1<\alpha<2, \\ (mn^{-1})^{1/2}N_n(x) \\ \quad +o_p\{\delta_m + (mn^{-1})^{1/2}\} & \text{if } 0<\alpha<1, \end{cases} \tag{2.15}$$

where the distribution of S_n converges to H, and $N_n(x)$ is asymptotically Normally distributed with zero mean and variance $\sigma(x)^2 > 0$.

Remark 2.4. Size of $\widetilde{H}_m - H_m$

When $1 < \alpha < 2$ the bootstrap involves centring at the sample mean, and the term $\widetilde{H}_m - H_m$ in (2.14) represents the dominant portion of the error arising from that part of the operation. By simple Taylor expansion, $\widetilde{H}_m - H_m \sim (mn^{-1})^{1-(1/\alpha)} H' S_n$, where $S_n = n^{-1/\alpha} \sum_{j \le n}(Y_j - \mu)$ converges in distribution to the stable law H as $n \to \infty$. Furthermore, $(c_n \log n)^{-1} S_n \to 0$ with probability one if and only if $\sum_{n \ge 2}(nc_n \log n)^{-1} < \infty$ (see Petrov (1975), pp.273-274). Therefore, $\widetilde{H}_m - H_m$ is of precise size $(mn^{-1})^{1-(1/\alpha)}$ in probability, and of smaller order than $(mn^{-1})^{1-(1/\alpha)}(\log n)^{\epsilon}$, for each $\epsilon > 0$, with probability 1. Furthermore, the terms in $(mn^{-1})^{2-(2/\alpha)}(\log n)^2 c_n^2$ and $(mn^{-1})^{2-(2/\alpha)}$ in (2.12) and (2.13) are of smaller order than $\widetilde{H}_m - H_m$.

When $0 < \alpha < 1$ the bootstrap does not involve any centring, and so $\widetilde{H}_m - H_m$ vanishes.

Remark 2.5. Size of M_m

We claim that M_m is of size $(mn^{-1})^{1/2}$ in probability and $(mn^{-1} \log\log n)^{1/2}$ almost surely. To appreciate why, note that for each x,

$$m^{-1} n E\{M_m(x)^2\} \to \sigma(x)^2 = b \int_{-\infty}^{\infty} \{H(x+y) - H(x)\}^2 |y|^{-(\alpha+1)} dy,$$

where $b > 0$ is such that $H'(y) \sim b|y|^{-(\alpha+1)}$ as $|y| \to \infty$; and that, by Lyapounov's central limit theorem, $(m^{-1}n)^{1/2} M_m(x)$ is asymptotically Normal $N\{0, \sigma(x)\}$. It is readily proven from Bernstein's inequality that $M_m(x) = O\{(mn^{-1} \log n)^{1/2}\}$ with probability one, and a longer argument may be employed to show that $M_m(x) = O\{(mn^{-1} \log\log n)^{1/2}\}$. (This requires, in addition to the conditions of Thorem 2.2, the assumption that $m(n)$ is nondecreasing and $m(n+1) - m(n) = O(1)$.)

3. Numerical Results

We took $Y = (\operatorname{sgn} X)|X|^{-1/\alpha}$, where X was Normal $N(0,1)$ and $1 < \alpha < 2$. In this setting, $a = (\pi/2)^{1/2} / \{\Gamma(\alpha) \sin(\alpha\pi/2)\}$, and the distribution function H of the stable law with characteristic function $\exp(-a|t|^{\alpha})$ may be calculated by numerical integration from the formula

$$H(x|\alpha, a) = \frac{1}{2} + \pi^{-1} \int_0^{\infty} t^{-1} \sin(tx) \exp(-at^{\alpha}) dt.$$

An asymptotic approximation to the distribution function $J_n(x) = G_n(n^{-1/\alpha}x) = P\{\sum_{j \le n}(Y_j - \mu) \le x\}$ is furnished by $\hat{J}_{1n} = H(n^{-1/\hat{\alpha}} x | \hat{\alpha}, \hat{a})$, where $\hat{\alpha}$ and $\hat{a}$

are defined using the method of Hill (1975). Our main aim is to compare this approach with the subsample bootstrap, which produces the estimator $\hat{J}_{2,n}(x) = \widehat{G}_m(n^{-1/\alpha}x)$. A third approximation is via the mixture distribution at (2.2), which suggests $\hat{J}_{3,n}(x) = \widetilde{G}_m(n^{-1/\alpha}x)$. (We replaced $H(n^{-1/\alpha}x|\hat{\alpha},\hat{a})$ by $H(n^{-1/\hat{\alpha}}x|\hat{\alpha},\hat{a})$, to reflect the fact that α would be unknown in paractical applications of this method.) In each case we make the comparison using the uniform metric to measure distance: $\hat{d}_j = \sup|\hat{J}_{j,n} - J_n|$.

Let $|Y_{(1)} - \bar{Y}| \leq \cdots \leq |Y_{(n)} - \bar{Y}|$ denote the ordered values of $|Y_i - \bar{Y}|$. To implement Hill's approach we put

$$\hat{\alpha} = (k+1)\Big(\sum_{j=1}^{k} \log|Y_{(n-j+1)} - \bar{Y}| - k\log|Y_{(n-k)} - \bar{Y}|\Big)^{-1},$$

$$\hat{a} = \widehat{C}\pi/\{\Gamma(\hat{\alpha})\sin(\hat{\alpha}\pi/2)\}, \text{where } \widehat{C} = |Y_{(n-k)} - \bar{Y}|^{\hat{\alpha}}(k+1)/n.$$

Both the asymptotic and bootstrap methods involve selecting tuning parameters, the number of order statistics, k, and the subsample size, m respectively. To make our comparative study fair we conducted extensive simulations to determine, for each α, the optimal values of k and m in the sense of minimising $\hat{d}_j$.

Table 3.1 gives the average, over five samples (four in the case of the hybrid method), of $\hat{d}_j$ for $j = 1, 2, 3$, as a function of $\alpha = 1.1(0.1)1.9$, for sample sizes $n = 20, 50, 100$, and employing optimal tuning parameters. (Computation was very time-consuming, hence our restriction to only four or five samples.) In all cases the bootstrap performs consistently worse than the asymptotic approach, as predicted by our theory. However, the hybrid method initially (for $n = 20$) performs worse than the asymptotic approach. By $n = 50$ our large-sample theory is more appropriate; there, the hybrid method outperforms both the bootstrap and the asymptotic approaches for the majority of values of α. By $n = 100$ the hybrid method is definitely outperforming both the bootstrap and the asymptotic approach.

Table 3.1. Approximations to $E(\hat{d}_j)$ for asymptotic, subsample bootstrap and hybrid methods

	α	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
	Asymptotic	0.063	0.094	0.148	0.062	0.082	0.063	0.063	0.062	0.064
$n = 20$	Bootstrap	0.181	0.149	0.216	0.130	0.160	0.113	0.110	0.101	0.093
	Hybrid	0.095	0.102	0.185	0.100	0.146	0.094	0.097	0.100	0.093
	Asymptotic	0.096	0.088	0.074	0.065	0.076	0.071	0.059	0.056	0.057
$n = 50$	Bootstrap	0.138	0.103	0.112	0.115	0.083	0.080	0.086	0.070	0.079
	Hybrid	0.065	0.058	0.061	0.050	0.049	0.061	0.075	0.065	0.068
	Asymptoctic	0.090	0.073	0.063	0.063	0.063	0.060	0.054	0.058	0.049
$n = 100$	Bootstrap	0.118	0.145	0.100	0.099	0.080	0.108	0.071	0.061	0.085
	Hybrid	0.048	0.042	0.047	0.046	0.027	0.040	0.050	0.043	0.073

The α values where the hybrid method performed worst were generally the larger ones. To some extent this is predicted by our theory, in that the convergence rates of all approaches have been shown to be poor for large α, being worse than $n^{-\epsilon}$ for any given $\epsilon > 0$ if α is sufficiently close to 2. This fact, coupled with the variability of particularly the bootstrap method (see below), indicates that any of the three methods may tend to outperform the others for such α's. However, we did not notice the predicted inferior performance of the hybrid method for α close to 1. In general, the hybrid method seems to perform well for moderate to large sample sizes.

The variability of any of the different methods may be measured by the average (over the nine values of α used in Table 3.1) of the sample variances (for the four or five different realizations) of the value of d_j corresponding to that method. For each of the three sample sizes considered, this measure shows that the bootstrap is the most variable, followed by the hybrid method. The same result is immediately seen from a plot of the five curves corresponding to values (for a given sample) of d_j as a function of α. Consistently, for each of the three sample sizes, the bootstrap method produces more variable curves than the hybrid method.

4. Proofs

Proof of Theorem 2.1. Suppose first that $1 < \alpha \leq 4/3$. In view of the smoothness conditions imposed on F, the functions $S(x) = P(|Y| > x) - 2F'(0)x^{-\alpha}$ and $D(x) = P(Y > x) - P(Y < -x) - F''(0)x^{-2\alpha}$ satisfy $|S(x)| + |D(x)| = O(x^{-(2\alpha+\eta)})$ as $x \to \infty$, for some $\eta > 0$. Therefore, as $t \downarrow 0$,

$$\int_0^\infty (x - \sin x)S(x/t)dx = O(t^{2\alpha+\eta}), \int_0^\infty (1 - \cos x)D(x/t)dx = O(t^{2\alpha+\eta}).$$

Hence, for $\eta > 0$ sufficiently small,

$$\begin{aligned} 1 - E(\cos tY) &= t\int_o^\infty (\sin tx)\{2F'(0)x^{-\alpha} + S(x)\}dx \\ &= a_1 t^\alpha + a_2' t^2 + O(t^{2\alpha+\eta}), \qquad (4.1) \\ E(\sin tY) &= \mu t - t\int_0^\infty (1 - \cos tx)\{F''(0)x^{-2\alpha} + D(x)\}dx \\ &= \mu t - a_3 t^{2\alpha} + O(t^{2\alpha+\eta}), \qquad (4.2) \\ \psi(t) = E(e^{itZ}) &= 1 - \{a_1|t|^\alpha - ia_1\mu(\operatorname{sgn} t)|t|^{\alpha+1} + a_2 t^2 \\ &\quad + ia_3(\operatorname{sgn} t)|t|^{2\alpha}\} + O(t^{2\alpha+\eta}). \qquad (4.3) \end{aligned}$$

Assume next that $4/3 < \alpha < 2$. Define $S(x)$ as before, but now put $D(x) = P(Y > x) - P(Y < -x)$. The analogues of (4.1)-(4.3) are here,

$$1 - E(\cos tY) = a_1|t|^\alpha + a_2' t^2 + O(|t|^{4-\alpha+\eta}),\ E(\sin tY) = \mu t + O(|t|^{4-\alpha+\eta}),$$
$$\psi(t) = 1 - \{a_1|t|^\alpha - ia_1\mu(\operatorname{sgn} t)|t|^{\alpha+1} + a_2 t^2\} + O(|t|^{4-\alpha+\eta}).$$

Henceforth we treat only the case $1 < \alpha \le 4/3$, bearing in mind that the last-written formula should be used in place of (4.3) when $\alpha > 4/3$.

Put $\epsilon = \eta/\alpha$, $z_1 = a_1|t|^\alpha + n^{1-(2/\alpha)}a_2t^2$, $z_2 = n^{-1/\alpha}ia_1\mu(\operatorname{sgn} t)|t|^{\alpha+1} - n^{-1}ia_3(\operatorname{sgn} t)|t|^{2\alpha}$, $z_3 = n^{-(1+\epsilon)}(|t| + t^{10})$. Then for $|t| \le n^\xi$ and $\xi > 0$ sufficiently small,

$$\begin{aligned}\{\psi(t/n^{1/\alpha})\}^n &= [1 - n^{-1}\{z_1 - z_2 + O(z_3)\}]^n\\ &= \exp(-a_1|t|^\alpha)\Big[1 + \sum_{j=1}^{2}\frac{1}{j!}(-n^{1-(2/\alpha)}a_2t^2)^j\\ &\quad + z_2 - n^{-1}\frac{1}{2}a_1^2|t|^{2\alpha} + O\{z_3 + (n^{1-(2/\alpha)}t^2)^3\}\Big]. \qquad (4.4)\end{aligned}$$

Condition (2.5) implies that the distribution of Z satisfies Cramér's conditions: $\sup_{|t|>\epsilon}|\psi(t)| < 1$ for all $\epsilon > 0$. From that condition and (2.5) we may deduce the existence of a constant $C \in (0,1)$ such that $|\psi(t)| < 1 - C\min(1, |t|^\alpha)$ for all real t. Hence, for any $\kappa > \xi > 0$,

$$\Big(\int_{n^\xi}^{n^\kappa} + \int_{-n^\kappa}^{-n^\xi}\Big)|\psi(t/n^{1/\alpha})|^n \le 2\int_{n^\xi}^{n^\kappa}\exp\{-C\min(t^\alpha, n)\}dt = O(n^{-K})$$

for all $K > 0$. It is straightforward to prove that, with

$$\begin{aligned}\pi(t) = \sum_{j=1}^{K}\frac{1}{j!}(-n^{1-(1/\alpha)}a_2t^2)^j + n^{-1/\alpha}ia_1\mu(\operatorname{sgn} t)|t|^{\alpha+1}\\ -n^{-1}(\frac{1}{2}a_1^2 + ia_3\operatorname{sgn} t)|t|^{2\alpha}, \qquad (4.5)\end{aligned}$$

we have

$$\Big(\int_{n^\xi}^{n^\kappa} + \int_{-n^\kappa}^{-n^\xi}\Big)\exp(-a_1|t|^\alpha)|\pi(t)|dt = O(n^{-K}).$$

Therefore,

$$\Big(\int_{n^\xi}^{n^\kappa} + \int_{-n^\kappa}^{-n^\xi}\Big)|\{\psi(t/n^{1/\alpha})\}^n - \exp(-a_1|t|^\alpha)\pi(t)|dt = O(n^{-K})$$

for all $K > 0$. Combining this result and (4.4), and remembering that the latter is valid for $|t| \le n^\xi$, we deduce that for a $\kappa > 0$,

$$\int_0^{n^\kappa}|t|^{-1}|\{\psi(t/n^{1/\alpha})\}^n - \exp(-a_1|t|^\alpha)\pi(t)|dt = o(n^{-1} + n^{2-(4/\alpha)}).$$

The theorem follows from this formula and the smoothing lemma for characteristic functions (Petrov (1975), Theorem 2, p.109).

Proof of (2.12). In the case $1 < \alpha < 2$ we may assume without loss of generality that $\mu = E(Y) = 0$, and define $Z_j = Y_j$ for both $0 < \alpha < 1$ and $1 < a < 2$. Put

$W = 0$ if $0 < \alpha < 1$ and $W = m^{1-(1/\alpha)} n^{-1} \sum_{j \le n} Z_j$ if $1 < \alpha < 2$. Let $\widehat{\psi}(t) = n^{-1} \sum_{j \le n} \exp(itZ_j)$ denote the empirical characteristic function of the data $\mathcal{X}$. Then $\psi = E(\widehat{\psi})$ denotes the characterisitc function of Z, and $\widehat{\psi}(t/m^{1/\alpha})^m e^{-itW}$ is the characteristic function of $S_m^* = m^{-1/\alpha} \sum_{j \le m} Z_j^*$, conditional on $\mathcal{X}$. Put $\Delta = \widehat{\psi} - \psi$ and $D(t) = m|\Delta(t/m^{1/\alpha})|$, let λ_n denote a sequence of positive numbers diverging so fast that $\lambda_n / \log n \to \infty$ yet so slowly that $mn^{-1}\lambda_n \log n \to 0$, and put $t_n = \min\{(n/m\lambda_n)^{1/\alpha}, m^{1/\alpha}\}$.

Our proof of the theorem is in five steps.

Step (a). There exists a set $\mathcal{E}_1 = \mathcal{E}_1(n)$, satisfying $P(\mathcal{E}_1) = 1 - O(n^{-K})$ for all $K > 0$, on which for constants C_1, $C_2 > 0$,

$$\begin{aligned} &|\widehat{\psi}(t/m^{1/\alpha})^m - \psi(t/m^{1/\alpha})^m - m\Delta(t/m^{1/\alpha})\psi(t/m^{1/\alpha})^{m-1}| \\ &\le C_1 D(t)^2 \exp(-C_2|t|^\alpha) \end{aligned} \tag{4.6}$$

uniformly in $|t| \le t_n$.

To derive (4.6) observe that by the binomial theorem,

$$\begin{aligned} |\widehat{\psi}^m - \psi^m - m\Delta\psi^{m-1}| &\le \sum_{j=2}^{m} \binom{m}{j} |\Delta|^j |\psi|^{m-j} \\ &\le \frac{1}{2} m^2 |\Delta|^2 |\psi|^m \exp\{(m-2)|\Delta/\psi|\}. \end{aligned} \tag{4.7}$$

Let $C_2 > 0$ be such that $|\psi(t)| \le \exp(-C_2|t|^\alpha)$ for $|t| \le 1$, and put $C_3^{-1} = \inf_{|t| \le \zeta} |\phi(t)|$, where $\zeta > 0$ is chosen sufficiently small for the infimum to be nonzero. Then, when the argument of Δ and ψ is $t/m^{1/\alpha}$ and $|t| \le \zeta m^{1/\alpha}$,

$$|\psi|^m \exp(|m\Delta/\psi|) \le \exp\{C_3 m|\Delta(t/m^{1/\alpha})| - C_2|t|^\alpha\}. \tag{4.8}$$

In Step (d) we shall derive the following result.

Lemma 4.1. *For all $K > 0$,*

$$\sup_{|t| \le 1} P[|\widehat{\psi}(t) - \psi(t)| > \max\{(n^{-1}\lambda_n|t|^\alpha)^{1/2}, n^{-1}\lambda_n\}] = O(n^{-K}).$$

In view of the lemma, $\sup_{|t| \le t_n} P\{D(t) > 1\} = O(n^{-K})$ for all $K > 0$. It follows that if $\mathcal{A}_n$ is a set of t values satisfying $|t| \le t_n$, and if $\mathcal{A}_n$ contains no more than $O(n^A)$ elements for an arbitrary but fixed $A > 0$, then

$$P\Big\{\sup_{t \in \mathcal{A}_n} D(t) > 1\Big\} \le \sum_{t \in \mathcal{A}_n} P\{D(t) > 1\} = O(n^{-K}) \tag{4.9}$$

for all $K > 0$. If t_1, t_2 are real numbers then

$$|D(t_1) - D(t_2)| \le C_4 m^{1-(1/\alpha)} |t_1 - t_2| n^{-1} \sum_{i=1}^{n} |Z_i| + |\psi(t_1) - \psi(t_2)|. \tag{4.10}$$

By choosing $\mathcal{A}_n$ to be a lattice with points n^{-B} apart, for arbitrarily large but fixed B, we may deduce from (4.9) and (4.10) that the event $\mathcal{E}_1 = \{\sup_{|t|\le t_n} D(t) \le 2\}$ satisfies $P(\mathcal{E}_1) = 1 - O(n^{-K})$ for all $K > 0$. Combining (4.7) and (4.8) we deduce that on $\mathcal{E}_1$,

$$|\widehat{\psi}(t/m^{1/\alpha})^m - \psi(t/m^{1/\alpha})^m - m\Delta(t/m^{1/\alpha})\psi(t/m^{1/\alpha})^{m-1}| \\ \le \frac{1}{2}D(t)^2 \exp(C_3 - C_2|t|^\alpha),$$

which gives (4.6) with $C_1 = \frac{1}{2}\exp(C_3)$.

Step (b). With probability one,

$$I_1 \equiv \int_{|t|\le t_n} |t|^{-1} D(t)^2 \exp(-C_2|t|^\alpha)dt = O(m\lambda_n/n). \tag{4.11}$$

To derive (4.11), note that by Lemma 4.1,

$$\sup_{|t|\le m^{1/\alpha}} P[D(t) > \max\{(mn^{-1}\lambda_n|t|^\alpha)^{1/2}, mn^{-1}\lambda_n\}] = O(n^{-K})$$

for all $K > 0$. The argument following (4.9) may now be used to show that the event

$$\mathcal{E}_2 = \Big\{ \sup_{|t|\le m^{1/\alpha}} D(t)/\max[(mn^{-1}\lambda_n|t|^\alpha)^{1/2}, mn^{-1}\lambda_n] \le 1 \Big\}$$

satisfies $P(\mathcal{E}_2) = 1 - O(n^{-K})$. On $\mathcal{E}_2$,

$$\begin{aligned} I_2 &\equiv \int_{(mn^{-1}\lambda_n)^{1/\alpha}\le|t|\le t_n} |t|^{-1} D(t)^2 \exp(-C_2|t|^\alpha)dt \\ &\le 2\int_{(mn^{-1}\lambda_n)^{1/\alpha}\le|t|\le t_n} t^{-1}\max\{mn^{-1}\lambda_n t^\alpha, (mn^{-1}\lambda_n)^2\}\exp(-C_2 t^\alpha)dt \\ &\le 2mn^{-1}\lambda_n \int_0^\infty t^{\alpha-1}\exp(-C_2 t^\alpha)dt = C_4 m\lambda_n/n, \end{aligned}$$

say. By the Borel-Cantelli lemma, $P(\tilde{\mathcal{E}}_2 \text{ i.o.}) = 0$, where $\tilde{\mathcal{E}}_2$ denotes the complement of $\mathcal{E}_2$. Therefore, with probability one,

$$I_2 = O(m\lambda_n/n). \tag{4.12}$$

Finally, we show that

$$I_3 \equiv \int_{|t|\le(mn^{-1}\lambda_n)^{1/\alpha}} |t|^{-1} D(t)^2 \exp(-C_2|t|^\alpha)dt = o(m\lambda_n/n), \tag{4.13}$$

with probability one.

The method of proof involves first using standard bounds for $|\sin x|$, $|\sin x - x|$ and $1-\cos x$ to prove that, with $\eta = \eta(n) = (\lambda_n/n)^{1/2}$,

$$I_3 \le C_5 m^2 \Big[\int_{|t|\le\eta} |t|^{-1} \{n^{-1}\sum_{i=1}^n \min(1,|tZ_i|)\}^2 dt + \eta^{2\alpha}\Big]$$

for $0<\alpha<1$, and

$$I_3 \le C_5 m^2 \Big\{ \int_{|t|\le\eta} |t|^{-1} \Big([n^{-1}\sum_{i=1}^n \min\{1,(tZ_i)^2\}]^2 + \{n^{-1}\sum_{i=1}^n \min(|tZ_i|,|tZ_i|^3)\}^2$$
$$+\{n^{-1}\sum_{i=1}^n \min(1,|tZ_i|) I(|Z_i|>\eta^{-1})\}^2\Big) dt + \eta^2 \Big| n^{-1}\sum_{i=1}^n Z_i I(|Z_i|\le\eta^{-1})\Big|^2 + \eta^{2\alpha}\Big\}$$

for $1<\alpha<2$. Expanding the squares of series as double series, and integrating the latter term-by-term, we may prove that $I_3 = O\{m^2 T_\alpha^2 + \eta^{2\alpha} + (\log n) U_3^2\}$, where $T_\alpha = \eta U_1$ if $0<\alpha<1$, $T_\alpha = \eta^2 U_2 + \eta|U_4|$ if $1<\alpha<2$,

$$U_1 = n^{-1}\sum_{i=1}^n |Z_i| I(|Z_i|\le\eta^{-1}) \quad \text{for } 0<\alpha<1,$$
$$U_2 = n^{-1}\sum_{i=1}^n Z_i^2 I(|Z_i|\le\eta^{-1}) \quad \text{for } 1<\alpha<2,$$
$$U_3 = n^{-1}\sum_{i=1}^n I(|Z_i|>\eta^{-1}) \quad \text{for } 0<\alpha<1 \text{ or } 1<\alpha<2,$$
$$U_4 = n^{-1}\sum_{i=1}^n Z_i I(|Z_i|\le\eta^{-1}) \quad \text{for } 1<\alpha<2.$$

Define $u_1 = u_4 = \eta^{\alpha-1}$, $u_2 = \eta^{\alpha-2}$ and $u_3 = \eta^\alpha$, and put $p_i = P\{|U_i - E(U_i)| > u_i\}$. It may be proved that $|E(U_i)| = O(u_i)$, and from Bernstein's inequality that $p_i = O\{\exp(-C_6 n\eta^\alpha)\} = O\{\exp(-C_6\lambda_n)\}$. Therefore, since $\lambda_n \to \infty$ faster than $\log n$, the Borel-Cantelli lemma implies that $|U_i| = O(u_i)$ with probability 1, for $1\le i\le 4$. Hence, $I_3 = O(m^2\eta^{2\alpha}\log n) = o(mn^{-1}\log n)$. This completes the proof of (4.13).

Step (c). Completion

Put $t_n' = \min(t_n, m^\xi)$. By assumption,

$$\int_{|t|\le t_n'} |t|^{-1} |\psi(t/m^{1/\alpha})^m - \exp(-a|t|^\alpha)\pi_m(t)| dt = o(\delta_m). \tag{4.14}$$

We claim that

$$\int_{|t|\le t_n'} |t|^{-1} m |\Delta(t/m^{1/\alpha})\{\psi(t/m^{1/\alpha})^{m-1} - \exp(-a|t|^\alpha)\pi_m(t)\}| dt$$
$$= o(\delta_m) + O\{(\lambda_n/mn)^{1/2}\} \tag{4.15}$$

with probability one. To appreciate why, note that the left-hand side is dominated by

$$\Big\{\int_{|t|\le t_n'} |t|^{-1}D(t)^2\exp(-C_7|t|^\alpha)dt\Big\}^{1/2}$$
$$\times\Big\{\int_{|t|\le t_n'} |t|^{-1}|\psi(1/m^{1/\alpha})^{m-1}-\exp(-a_1|t|^\alpha)\pi_m(t)|^2\exp(C_7|t|^\alpha)dt\Big\}^{1/2},$$

where $C_7>0$ is arbitrary. The arguments in Step(b) show that the first integral in the expression equals $O(m\lambda_n/n)=o(1)$, with probability one. The second integral is dominated by $C_8(I_4+I_5)$, where

$$I_4=\int_{|t|\le t_n'} |t|^{-1}|\psi(t/m^{1/\alpha})^m-\exp(-a|t|^\alpha)\pi_m(t)|^2\exp(C_7|t|^\alpha)dt=o(\delta_m^2),$$
$$I_5=\int_{|t|\le t_n'} |t|^{-1}|\psi(t/m^{1/\alpha})|^{2(m-1)}|1-\psi(t/m^{1/\alpha})|^2\ \exp(C_7|t|^\alpha)dt$$
$$\le C_9\int_{|t|\le t_n'} |t|^{-1}\exp(-C_{10}|t|^\alpha)|t/m^{1/\alpha}|^{2\alpha}\exp(C_7|t|^\alpha)dt=O(m^{-2}),$$

provided C_7 is sufficiently small. Result (4.15) is immediate.

Formulae (4.6) and (4.11) imply that

$$\int_{|t|\le t_n'}|t|^{-1}|\hat\psi(t/m^{1/\alpha})^m-\psi(t/m^{1/\alpha})^m-m\Delta(t/m^{1/\alpha})\psi(t/m^{1/\alpha})^{m-1}|dt=O(m\lambda_n/n)$$

with probability one. This property, (4.14) and (4.15) together yield that

$$\int_{|t|\le t_n'} |t|^{-1}|\widehat\psi(t/m^{1/\alpha})^m-\{1+m\Delta(t/m^{1/\alpha})\}\exp(-a_1|t|^\alpha)\pi_m(t)|dt$$
$$=o(\delta_m)+O\{(m\lambda_n/n)+(\lambda_n/mn)^{1/2}\} \tag{4.16}$$

with probability one.

The following lemma is proved in Step (e).

Lemma 4.2. *There exist constants $C_{11},C_{12}>0$ such that for all $1\le m\le n$ and all real t, $E|\widehat\psi(t)|^m\le C_{11}\exp\{-mC_{12}(|t|^\alpha\wedge 1)\}$.*

Define

$$I_4=\int_{t_n'<|t|\le n^\kappa} |t|^{-1}E|\widehat\psi(t/m^{1/\alpha})|^m dt,$$

where $\kappa>1$ is fixed but arbitrarily large. In view of Lemma 4.2,

$$E(I_4)\le C_{13}\int_{t_n'}^{n^\kappa} t^{-1}\exp\{-C_{12}(|t|^\alpha\wedge 1)\}dt\le C_{14}n^\kappa\exp[-C_{12}\{(n/m\lambda_n)\wedge m^{(\alpha\xi)\wedge 1}\}].$$

By definition of λ_n, $(n/m\lambda_n)/\log n\to\infty$, and since m is an order of magnitude larger than each positive power of $\log n$, $m^{(\alpha\xi)\wedge 1}/\log n\to\infty$. Therefore, $E(I_4)=$

$O(n^{-K})$ for all $K > 0$. It now follows from the Borel-Cantelli lemma that for each $K > 0, I_4 = O(n^{-K})$ with probability one.

It is straightforward to show that with probability one,

$$\int_{t_n' < |t| \leq n^\kappa} |t|^{-1} |\{1 + m\Delta(t/m^{1/\alpha})\} \exp(-a_1|t|^\alpha)\pi_m(t)| dt = O(n^{-K})$$

for all $K > 0$. Hence, with

$$A(t) = \widehat{\psi}(t/m^{1/\alpha})^m - \{1 + m\Delta(t/m^{1/\alpha})\} \exp(-a_1|t|^\alpha)\pi_m(t),$$

we have

$$\int_{t_n' < |t| \leq n^\kappa} |t^{-1} A(t)| dt = O(n^{-K}).$$

This result and (4.16) together imply that for all $\kappa > 0$,

$$\int_{|t| \leq n^\kappa} |t^{-1} A(t)| dt = o(\delta_m) + O\{m\lambda_n/n) + (\lambda_n/mn)^{1/2}\} \tag{4.17}$$

with probability one.

When $0 < \alpha < 1$, A equals the Fourier-Stieltjes transform of the function $\widehat{G}_m - (\widetilde{H}_m + M_m) = \widehat{G}_m - (H_m + M_m)$. Theorem 2.2 then follows from (4.17) via the smoothing lemma for characteristic functions (Petrov (1975), p.109). In the case $1 < \alpha < 2$ the Fourier-Stieltjes transform of $\widehat{G}_m - (\widetilde{H}_m + M_m)$ equals

$$B(t) = \widehat{\psi}(t/m^{1/\alpha})^m e^{-itW} - \{e^{-itW} + m\Delta(t/m^{1/\alpha})\} \exp(-a_1|t|^\alpha)\pi_m(t),$$

and

$$|B(t)| \leq |A(t)| + m|\Delta(t/m^{1/\alpha})\pi_m(t)||e^{itW} - 1| \exp(-a_1|t|^\alpha).$$

The argument leading to (4.11) may be used to prove that

$$\int_{|t| \leq n^\kappa} m|\Delta(t/m^{1/\alpha})\pi_m(t)| \exp(-a_1|t|^\alpha) dt = O\{(m\lambda_n/n)^{1/2}\}.$$

Therefore, since $|t^{-1}(e^{itW} - 1)| \leq 2|W| = O\{(m/n)^{1-(1/\alpha)}(c_n \log n)^{1/\alpha}\}$ with probability one (Petrov (1975), p.274),

$$\int_{|t| \leq n^\kappa} |t^{-1} B(t)| dt = o(\delta_m) + O\{(m\lambda_n/n) + (\lambda_n/mn)^{1/2} + (m/n)^{2-(2/\alpha)}(c_n \log n)^{2/\alpha}\}$$

with probability one. Theorem 2.2 follows from this result via the smoothing lemma for characteristic functions. (An argument by contradiction may be used to show that λ_n may be replaced by $\log n$; note that $\lambda_n / \log n \to \infty$ arbitrarily slowly.)

Step (d). Proof of Lemma 4.1.

Let $B_1, B_2, \ldots$ denote positive constants and let ψ and $\widehat{\psi}$ have argument t, which at this stage we do not insist satisfies $|t| \leq 1$. Put $\widehat{\psi}_1(t) = n^{-1} \sum_{j \leq n} \cos(tZ_j)$, $\widehat{\psi}_2(t) = n^{-1} \sum_{j \leq n} \sin(tZ_j), \psi_k = E(\widehat{\psi}_k)$. Now,

$$
\begin{aligned}
n\text{Var}\,(\widehat{\psi}_1) &= \frac{1}{2}\{1 + \psi_1(2t)\} - \psi_1(t)^2 \leq B_1(|t|^\alpha \wedge 1), \\
n\text{Var}\,(\widehat{\psi}_2) &= \frac{1}{2}\{1 - \psi_2(2t)\} - \psi_2(t)^2 \leq B_1(|t|^\alpha \wedge 1).
\end{aligned}
$$

Hence, by Bernstein's inequality, for each $x > 0$,

$$
P(|\widehat{\psi} - \psi| > x) \leq \sum_{k=1}^{2} P(|\widehat{\psi}_k - \psi_k| > \frac{1}{2}x) \leq 4 \exp[-nB_2 x^2 / \{(|t|^\alpha \wedge 1) + x\}]. \quad (4.18)
$$

Therefore, if $|t| \leq 1$,

$$
P(|\widehat{\psi} - \psi| > x) \leq 4 \exp(-nB_3 x^2 |t|^{-\alpha}) + 4 \exp(-nB_3 x).
$$

Now take $x = \max\{(n^{-1}\lambda_n |t|^\alpha)^{1/2}, n^{-1}\lambda_n\}$ to deduce the lemma.

Step (e). Proof of Lemma 4.2.

If $||\widehat{\psi}|^2 - |\psi|^2| \leq \frac{1}{2}(1 - |\psi|^2)$ then

$$
|\widehat{\psi}|^2 = 1 - (1 - |\psi|^2) + (|\widehat{\psi}|^2 - |\psi|^2) \leq 1 - \frac{1}{2}(1 - |\psi|^2) \leq \exp\{-2B_4(|t|^\alpha \wedge 1)\}.
$$

If $\||\widehat{\psi}|^2 - |\psi|^2| > \frac{1}{2}(1 - |\psi|^2)$ then $|\widehat{\psi} - \psi| > \frac{1}{4}(1 - |\psi|^2) \geq B_5(|t|^\alpha \wedge 1)$. Therefore, using (4.18).

$$
\begin{aligned}
E|\widehat{\psi}|^m &\leq \exp\{-mB_4(|t|^\alpha \wedge 1)\} + P\{|\widehat{\psi} - \psi| > B_5(|t|^\alpha \wedge 1)\} \\
&\leq 5 \exp\{-mB_6(|t|^\alpha \wedge 1)\}.
\end{aligned}
$$

References

Arcones, M. A. and Giné, E. (1989). The bootstrap of the mean with arbitrary sample size. *Ann. Inst. H. Poincaré* **25**, 457-481.

Arcones, M. A. and Giné, E (1992). Additions and corrections to the bootstrap of the mean with arbitary sample size. *Ann. Inst. H. Poincaré* **27**, 583-595.

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

Beran, R. and Srivastava, M. S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.* **13**, 92-115.

Booth, J. G. and Hall. P. (1993). An improvement of the jackknife distribution function estimator. *Ann. Statist.* **21**, 1476-1485.

Bickel, P., Bötze, F. and Van Zwet, W. R. (1994). Resampling fewer than n observations: gains, losses, and remedies for losses. Tecxhnical Report No. 419, Department of Sttistics, University of California, Berkeley.

Bickel, P. and Yahav, J. A. (1988). Richardson extrapolation and the bootstrap. *J. Amer. Statist. Assoc.* **83**, 387-393.

Cramér, H. (1962). *On the Approximation to a Stable Probability Distribution.* Studies in Mathematical Analysis and Related Topics. Stanford Univrsity Press, Stnford.

Cramér, H. (1964). On asymptotic exansions for sums of independent random variables with limiting stble distribution. *Sankhyā* **25**, 13-24.

Csörgö, S., Deheuvels, P. and Mason, D. (1985). Kernel estimates of the tail index of a distribution. *Ann. Statist.* **13**, 1050-1077.

Deheuvels, P., Mason, D. M. and Shorack, G. R. (1993). Some results on the influence of extremes on the bootstrap. *Ann. Inst. H. Pointcaré* **29**, 83-103.

Giné, E. and Zinn, J. (1989). Necessary conditions for the bootstrap of the mean. *Ann. Statist.* **17**, 684-691.

Hall, P. (1981). On the rate of convergence to a stable law. *J. London Math. Soc.* **23**, 179-192.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. Roy. Statist. Soc. Ser. B* **44**, 37-42.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. (With discussion.) *Ann. Statist.* **16**, 927-985.

Hall, P. (1990a). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32**, 177-203.

Hall, P. (1990b). Asymptotic properties of the bootstrap for heavy-tailed distributions. *Ann. Probab.* **18**, 1342-1360.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **13**, 1163-1174.

Huang, J. S., Sen, P. K. and Shao, J. (1996). Bootstrapping a sample quantile when the dentsity has a jump. *Statist. Sinica* **6**, 299-309.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *Ann. Statist.* **17**, 1168-1175.

Mammen, E. (1992). *When does Bootstrap Work? Asymptotic Results and Simulations.* Lecture Notes in Statistics **77**. Springer-Verlag, New York.

Petrov, V. V. (1975). *Sums of Independent Random Variables.* Springer, New York.

Politis, D. N. and Romano, J. P. (1992). A general theory for large sample confidence regions based on subsamples under minimal assumptions. Technical Report No. 399, Department of Statistics, Stanford University.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031-2050.

Politis, D. N., Romano, J. P. and You, L. (1993). Uniform confidence bands for the spectrum based o subsamples. In *Computing Science and Statistics, Proceedings of the 25th Symposium on the Interface* (Edited by M. Tarter and M. Lock), 346-351. The Interface Foundation of North America.

Wu. C. F. J. (1990). On the asymptotic properties of the jackknife histogram. *Ann. Statist.* **18**, 1438-1452.

Zolotarev, V. M. (1962). An analogue of the asymptotic expansion of Edgeworth-Cramér. In Proc. 6th All-Union Conf. Probability Theory and Math. Statist. Vilnius, 1962.

Centre for Mathematics and Its Applications, Australian National University, Canberra, ACT 0200, Australia.

E-mail: Peter.Hall@anu.edu.au

E-mail: majing@ust.hk

(Received June 1996; accepted July 1997)