Ann. Inst. Statist. Math.
Vol. 48, No. 3, 403–421 (1996)

# ON BOOTSTRAP ESTIMATION OF THE DISTRIBUTION OF THE STUDENTIZED MEAN

PETER HALL[1] AND RAOUL LEPAGE[1,2]*

[1] *School of Mathematical Sciences, Centre for Mathematics and its Applications, Australian National University, Canberra, A.C.T. 0200, Australia*
[2] *Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, U.S.A.*

(Received July 29, 1994; revised October 20, 1995)

**Abstract.** It is shown that bootstrap methods for estimating the distribution of the Studentized mean produce consistent estimators in quite general contexts, demanding not a lot more than existence of finite mean. In particular, neither the sample mean (suitably normalized) nor the Studentized mean need converge in distribution. It is unnecessary to assume that the sampling distribution is in the domain of attraction of any limit law.

*Key words and phrases*: Bootstrap, central limit theorem, consistency, domain of attraction, domain of partial attraction, heavy tail, percentile-$t$ method, self-normalization, Stable law, Studentization.

## 1. Introduction

Among bootstrap methods for conducting inference about a mean, techniques based on self-normalization occupy a central position. In classical settings, where the tails of the sampling distribution are relatively light, self-normalization or Studentization has been discussed extensively in the context of the percentile-$t$ method, where pivotalness is the main issue. Self-normalization ensures second-order accuracy, and in the case of inference about a univariate mean it is arguably the simplest approach to accurate bootstrap inference. See for example Hall (1992). The self-normalized bootstrap has also been treated for relatively heavy-tailed distributions, where it has been shown to be appropriate when the sampling distribution comes from the domain of attraction of stable laws whose exponent exceeds 1. See in particular the work of Arcones and Giné (1989).

The cases of light- and heavy-tailed distributions differ considerably in the complexity of their technical prescriptions. The former demands only moment assumptions, but the latter requires intricate conditions on the tails of the sampling

---

* Now at Michigan State University.

distribution. Not only must the tail probabilities be regularly varying of an appropriate order, they must be balanced, in the sense that their ratios must enjoy a well-defined limit. In the present paper we show that conditions of this type are unnecessary, and that the attractiveness of the percentile-$t$ bootstrap extends to distributions that are not in the domain of attraction of any limit law. Indeed, bootstrap methods for the mean, based on self-normalization, are applicable under a condition that is not much stronger than existence of finite mean.

The remarks above apply to methods that are based on using a resample of smaller size than the sample, and which employ the bootstrap to approximate the distribution of the Studentized mean, $T_m$ say, for this smaller size, $m$. If we wish to approximate the distribution of the Studentized mean constructed for the original sample size, say $n$, then the distribution of $T_m$ must be close to that of $T_n$. This requires a more stringent assumption on the truncated second moment, but which is still substantially less than demanding that the sampling distribution be in a domain of attraction. Indeed, the distribution may be simultaneously in the domain of partial attraction of every Stable law (including that of the Normal law) whose exponent exceeds $1+\epsilon$ for some $\epsilon > 0$.

Related work includes that of Athreya (1987), Giné and Zinn (1989), Knight (1989), Hall (1990*a*) and Deheuvels *et al.* (1993), who analysed the influence that extreme summands have on bootstrap methods for the mean in the case of heavy-tailed distributions; Arcones and Giné (1991), who investigated similar issues, including consistency of both distribution and moment estimators based on the bootstrap; and Swanepoel (1986), Hall (1990*b*) and Wu (1990), who noted the importance of using resample sizes smaller than the sample size. Politis and Romano (1994) showed that the subsample bootstrap produces consistency in a wide range of settings. However, the fact that the statistics there require limiting distributions (see Politis and Romano's condition (A)) excludes much of the context studied in the present paper.

Section 2 presents our main theorem, whose implications and regularity conditions are addressed in Section 3. Examples describing aspects of the theorem are discussed in Section 4. Section 5 gives a proof of the theorem.

## 2. Main result

Let $X, X_1, X_2, \ldots$ denote independent and identically distributed random variables with finite mean $\mu$. Put $\bar{X} = n^{-1}\sum_{j=1}^n X_j$, the sample mean; $S^2 = n^{-1}\sum_{j=1}^n (X_j - \bar{X})^2$, the sample variance; and $T_n = n^{1/2}(\bar{X}-\mu)/S$, the centred, Studentized mean. The bootstrap version of $T_n$ may be introduced as follows. Conditional on $\mathcal{X} = \{X_1, \ldots, X_n\}$, let $X_1^*, \ldots, X_n^*$ denote independent and identically distributed random variables drawn randomly, with replacement, from $\mathcal{X}$; let $m \le n$; and put $\bar{X}_m^* = m^{-1}\sum_{j=1}^m X_j^*$, $S_m^{*2} = m^{-1}\sum_{j=1}^m (X_j^* - \bar{X}_m^*)^2$ and $T_m^* = m^{1/2}(\bar{X}_m^* - \bar{X})/S_m^*$. We claim that under very mild regularity conditions, including the assertion that $m = m(n) \to \infty$ and $m/n \to 0$, the conditional distribution of $T_m^*$, given the data $\mathcal{X}$, approximates the unconditional distribution of $T_m$; and that under slightly more stringent assumptions, the distributions of

$T_m$ and $T_n$ are close, so that the conditional distribution of $T_m^*$ approximates the unconditional distribution of $T_n$.

Our assumptions concerning the distribution of $X$ are expressed solely in terms of the sum and difference of the truncated covariance function,

$$\tau_\pm(x) = E\{X^2 I(0 < X \le x)\} \pm E\{X^2 I(0 < -X \le x)\}.$$

Put $\rho_\pm(x,\lambda) = \tau_\pm(\lambda x)/\tau_+(x)$ if $\tau_+(x) > 0$, 1 otherwise. We shall suppose that for some $\epsilon, C > 0$,

$$\lim_{\lambda\downarrow 1} \limsup_{x\to\infty} \rho_+(x,\lambda) = 1, \qquad \sup_{\lambda>1, x>C} \lambda^{\epsilon-1}\rho_+(x,\lambda) < \infty; \tag{2.1}$$

and on occasion that for all $0 < \zeta < 1$,

$$\sup_{\zeta x \le y \le x} |\rho_\pm(x,\lambda) - \rho_\pm(y,\lambda)| \to 0 \tag{2.2}$$

for each $\lambda > 0$, as $x \to \infty$.

THEOREM 2.1. *Assume* (2.1), *and that* $m = m(n) \to \infty$ *and* $m/n \to 0$. *Then*

$$\sup_{-\infty<x<\infty} |P(T_m \le x) - P(T_m^* \le x|\mathcal{X})| \to 0 \tag{2.3}$$

*in probability,*

$$\lim_{\epsilon\to 0, \lambda\to\infty} \liminf_{n\to\infty} P(\epsilon \le |T_n| \le \lambda) = 1 \tag{2.4}$$

*and*

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \sup_{-\infty<x<\infty} P(x < T_n \le x+\epsilon) = 0. \tag{2.5}$$

*If in addition* $mn^{-1}\log n \to 0$ *then the convergence in* (2.3) *is with probability one. If both* (2.1) *and* (2.2) *hold then there exists a sequence of positive constants* $\delta_n$ *decreasing to zero, such that*

$$\sup_{-\infty<x<\infty} |P(T_n \le x) - P(T_m \le x)| \to 0 \tag{2.6}$$

*whenever* $\delta_n \le m/n \to 0$.

## 3. Discussion

*Remark* 3.1. *Simpler, sufficient condition for* (2.1). A condition that is simpler than (2.1) and which implies that constraint is the following: for some $\epsilon > 0$ and each $\lambda_0 > 1$,

$$\tau_+(\lambda x) \leq \lambda^{1-\epsilon}\tau_+(x) \tag{3.1}$$

for all $\lambda \geq \lambda_0$ and all sufficiently large $x$. It is straightforward to check that conditions (2.2) and (3.1) both hold if the distribution of $X$ is in the domain of attraction of the Normal law, or of a Stable law whose exponent $\alpha$ exceeds 1. (The assumption $\alpha > 1$ ensures that $E(|X|) < \infty$. In this case the quantity $\epsilon$ in condition (3.1) may be taken to be any positive number less than $\alpha - 1$.) In such contexts $\delta_n$ may be any sequence such that $\delta_n + (n\delta_n)^{-1} \to 0$.

*Remark* 3.2. *Finiteness of moments.* Assumption (2.1) implies that $E(|X|^{1+\delta}) < \infty$ for all $0 < \delta < \epsilon$, and in particular that $E(|X|) < \infty$. To appreciate why, note that by (2.1) there exists a constant $C > 0$ such that

$$\begin{aligned} x^{1-\delta}E\{|X|^{1+\delta}I(|X| > x)\} &\leq \int_1^\infty u^{\delta-2}\tau_+(ux)du \\ &\leq C\tau_+(x)\int_1^\infty u^{\delta-1-\epsilon}du < \infty. \end{aligned}$$

The requirement that $X$ enjoy a property stronger than the existence of finite mean is not unexpected, since the conclusions of Theorem 2.1 fail if the sole assumption is that $X$ be in the domain of attraction of a Stable law with finite mean. Indeed, if $X$ has those properties and satisfies $P(X > C) = 1$ for a finite constant $C$ then it may be shown that $P(T_n > \lambda) \to 1$ as $n \to \infty$, for each $\lambda > 0$, and so (2.4) fails. This indicates that in such a context $T_n$ is inappropriate as a basis for inference about the mean.

*Remark* 3.3. *Distributions that are not in any domain of attraction.* Example 4.2 will discuss a distribution that satisfies (2.1) and (2.2), but is not in any domain of attraction. Moreover, it is in the domain of partial attraction of many stable laws. In such cases the distribution of $T_n$ gets arbitrarily close, along subsequences, to each of the limiting distributions of self-normalized sums whose sampling distributions come from the various domains of attraction, but the distribution of $T_n$ does not converge to anything along the full sequence of $n$'s.

In view of this curious and quite non-standard behaviour, it is important to know whether the distribution of $T_n$ is in some sense asymptotically bounded away from zero and infinity, and asymptotically continuous, despite the fact that it does not converge. Results (2.4) and (2.5) guarantee those properties. In particular, (2.5) ensures that in an asymptotic sense, the range of quantiles of the distribution of $T_n$ may be approximated arbitrarily closely by the continuum, and so bootstrap procedures are not confounded by the presence of large spikes of probability mass;

and (2.4) shows that no part of the distribution of $T_n$ "escapes to zero or infinity", ensuring that the bootstrap approximation described by (2.3) is not vacuous.

*Remark* 3.4. *Application to confidence procedures.* Together, (2.3) and (2.6) imply that

$$\sup_{-\infty<x<\infty} |P(T_n \le x) - P(T_m^* \le x \mid \mathcal{X})| \to 0, \tag{3.2}$$

which provides a bootstrap approximation to the distribution of $T_n$. Results (2.4), (2.5) and (3.2) ensure the consistency of the following bootstrap procedure for constructing a confidence interval with coverage probability $\pi \in (0,1)$. Put

$$\hat{x}_\pi = \sup\{x : P(|T_m^*| \le x \mid \mathcal{X}) \le \pi\},$$

and let $\mathcal{I}_\pi = (\bar{X} - n^{-1/2}\hat{x}_\pi S, \bar{X} + n^{-1/2}\hat{x}_\pi S)$. Then $\mathcal{I}_\pi$ is a nominal $\pi$-level confidence interval for $\mu$, and $P(\mu \in \mathcal{I}_\pi) \to \pi$ as $n \to \infty$. Of course, asymptotically valid inference may be conducted under (2.1) alone, without the need for (2.2), provided we base the confidence interval on $T_m$ rather than $T_n$. Indeed, let us define $\mathcal{I}'_\pi$ to be that version of $\mathcal{I}_\pi$ in which $n^{-1/2}$ is replaced by $m^{-1/2}$, and where $\bar{X}$ and $S$ are replaced by their counterparts based on the first $m$ elements of the sample $\mathcal{X}$. (The quantity $\hat{x}_\pi$ is left unaltered.) Then by (2.3)–(2.5), and provided $m$ is chosen so that $m^{-1} + mn^{-1} \to 0$, we have $P(\mu \in \mathcal{I}'_\pi) \to \pi$.

*Remark* 3.5. *Implications for empirical likelihood.* The technique of empirical likelihood (see Owen (1988, 1990)) is sometimes considered a competitor of the bootstrap. The "classical" form of empirical likelihood ratio methodology, in which the distribution of the log likelihood ratio statistic is compared with that of a chi-squared distribution, is not necessarily valid for heavy-tailed sampling distributions, since there the ratio is not asymptotically distributed as chi-squared. However, an alternative approach in which the distribution of the ratio is calibrated using the bootstrap is valid in the context of the sample mean computed from a heavy-tailed parent, provided that (2.1) holds and an appropriately smaller resample size is employed. Indeed, it may be shown that the log likelihood ratio evaluated at the true parameter value, $\mu$, is first-order equivalent to $T_n^2$, and that its bootstrap counterpart for a resample of size $m$ is first-order equivalent to $T_m^{*2}$. Therefore the results described in the previous paragraph justify using bootstrap-calibrated empirical likelihood (with a *resample* of size $m$) in cases where (2.1) holds, if the calibration is employed to adjust the distribution of the ratio computed for a *sample* of size $m$ rather than $n$; and justify its use under (2.1) and (2.2) if the calibration is applied to the ratio computed for a sample of size $n$.

## 4. Examples

*Example* 4.1. *Distribution in domain of attraction of stable law.* Let $X$ have density $f$, defined by

$$f(x) = \begin{cases} p\alpha x^{-\alpha-1} & \text{if} \quad x > 1 \\ (1-p)\alpha|x|^{-\alpha-1} & \text{if} \quad x < -1 \\ 0 & \text{if} \quad |x| \leq 1, \end{cases}$$

where $1 < \alpha < 2$ and $0 \leq p \leq 1$. Then $X$ is in the domain of attraction of a stable law with exponent $\alpha$ and tail balance ratio $p/(1-p)$. Conditions (2.1) and (2.2) hold, and $(n^{1-(1/\alpha)}(\bar{X}-\mu), n^{1-(2/\alpha)}S^2)$ converges in distribution to $(Y_1, Y_2)$, say, where $Y_1$ has a stable distribution with exponent $\alpha$ and $Y_2$ has a positive stable distribution with exponent $\alpha/2$. Therefore, $T_n$ converges in distribution to $Z = Y_1/Y_2^{1/2}$. The bootstrap approximation to the distribution of $T_n$ may be expressed as follows: if $m = m(n) \to \infty$ and $m/n \to 0$ then

$$P(T_m^* \leq x \mid \mathcal{X}) \to P(Z \leq x)$$

in probability, uniformly in $x$.

*Example* 4.2. *Distribution in domain of partial attraction of many stable laws.* This example is of a distribution in the domain of partial attraction of all symmetric stable laws with exponent $\alpha \in (1+\epsilon, 2-\epsilon)$ for any $\epsilon \in (0, 1/2)$. A more elaborate version of the same construction will produce a distribution in the domain of partial attraction of all stable laws with exponent $\alpha \in (1, 2]$ and tail balance ratio $p/(1-p) \in [0, \infty]$.

Given $\epsilon, \delta \in (0, 1/2)$ let $1 = x_1 < x_2 < \cdots$ be defined by $x_i = \prod_{1\leq j\leq i} j$, and let $\alpha_1, \alpha_2, \ldots$ be a dense subset of $(1+\epsilon, 2-\epsilon)$ with the property that for $i \geq 2$, $|\alpha_i - \alpha_{i+1}| \leq \delta/\{i(\log i)\log\log(i+1)\}$. (Selection of a dense sequence with this property is possible because $\sum\{i(\log i)\log\log(i+1)\}^{-1} = \infty$.) Note that $\log x_i \sim i \log i$, and put

$$c_i = \{(x_i^{-\alpha_i} + x_i^{-\alpha_{i+1}}) - (x_{i+1}^{-\alpha_{i+1}} + x_{i+1}^{-\alpha_{i+2}})\}/\{4(x_i^{-\alpha_i} - x_{i+1}^{-\alpha_i})\}.$$

By choosing $\delta$ sufficiently small we may ensure that each $c_i > 0$ and $c_i \to 1$ as $i \to \infty$. (To derive the latter result, note that $|\alpha_i - \alpha_{i+1}| \log x_i \to 0$ and

$$\begin{aligned} &x_i^{\max(\alpha_i,\alpha_{i+1})}/x_{i+1}^{\min(\alpha_i,\alpha_{i+1},\alpha_{i+2})} \\ &\quad \sim (x_i/x_{i+1})^{\alpha_i} \exp[O\{|\alpha_i - \alpha_{i+1}| \log x_i \\ &\qquad\qquad + (|\alpha_i - \alpha_{i+1}| + |\alpha_{i+1} - \alpha_{i+2}|) \log x_{i+1}\}] \\ &\quad \sim (x_i/x_{i+1})^{\alpha_i} \to 0.) \end{aligned}$$

Define

$$f(x) = \begin{cases} 0 & \text{if} \quad |x| \leq x_1 \\ c_i \alpha_i |x|^{-\alpha_i - 1} & \text{if} \quad x_i < |x| \leq x_{i+1}, \text{ for } i \geq 1. \end{cases}$$

The definition of $c_i$ ensures that $\int f = 1$ and $f \geq 0$. Let $X$ have density $f$. We shall prove that this distribution satisfies conditions (2.1) and (2.2). Similarly it may be shown that $X$ is in the domain of partial attraction of every symmetric stable law with exponent between $1+\epsilon$ and $2-\epsilon$. Clearly, from the tail properties of the distribution of $X$, it is not in any domain of attraction.

Define $\alpha(x) = \alpha_i$ if $\alpha_i < x \leq x_{i+1}$. As $m \to \infty$,

$$\begin{aligned}(4.1) \qquad \int_{|x|\leq x_{m+1}} x^2 f(x)dx &= 2\sum_{i=1}^{m} c_i\alpha_i(2-\alpha_i)^{-1}(x_{i+1}^{2-\alpha_i} - x_i^{2-\alpha_i}) \\ &\sim \alpha_m(2-\alpha_m)^{-1}x_{m+1}^{2-\alpha_m}.\end{aligned}$$

Hence, there exist constants $C_1, C_2 > 0$ such that

$$(4.2) \qquad C_1 x^{2-\alpha(x)} \leq E\{X^2 I(|X| \leq x)\} \leq C_2 x^{2-\alpha(x)}$$

uniformly in $x > x_2$. Result (4.1), and the properties of $\alpha_i$, imply that for each $0 < \zeta < 1$ and $0 < \lambda_1 < \lambda_2 < \infty$,

$$\sup_{\zeta x < y \leq x, \lambda_1 \leq \lambda \leq \lambda_2} |\rho_+(y,\lambda) - \lambda^{2-\alpha(x)}| \to 0$$

as $x \to \infty$. This is enough to give the first part of condition (2.1), and also (2.2) in the case of the + sign. Condition (2.2) in the case of the − sign is trivial, because $X$ has a symmetric distribution. Since $|\alpha(\lambda x) - \alpha(x)| \log x \to 0$ as $x \to \infty$, uniformly in $1 \leq \lambda \leq x^C$ for each fixed $C > 0$, then by (4.2), for all $x \geq x_2$ and $1 \leq \lambda \leq x^{2/\epsilon}$,

$$\begin{aligned}(4.3) \qquad \rho_+(x,\lambda) &\leq C_1^{-1}C_2(\lambda x)^{2-\alpha(\lambda x)}x^{\alpha(x)-2} \\ &\leq C_3\lambda^{2-\alpha(x)} \leq C_3\lambda^{1-\epsilon}.\end{aligned}$$

If $\lambda > x^{2/\epsilon}$ then $x \leq \lambda^{\epsilon/2}$, and so

$$\begin{aligned}(4.4) \qquad \rho_+(x,\lambda) &\leq C_1^{-1}C_2(\lambda^{1+(\epsilon/2)})^{2-\alpha(x)} \\ &\leq C_1^{-1}C_2(\lambda^{1+(\epsilon/2)})^{1-\epsilon} \leq C_1^{-1}C_2\lambda^{1-(\epsilon/2)}.\end{aligned}$$

Results (4.3) and (4.4) imply the second part of (2.1).

*Example* 4.3. *Smooth functions of vectors of means*. Here we consider the case of a statistic $\hat{\theta}$ which can be expressed as a smooth function of a vector of sample means. Assume that the $r$-vectors $X_j = (X_j^{(1)}, \ldots, X_j^{(r)})$, $1 \leq j \leq n$, are independent and identically distributed with mean $\mu = (\mu^{(1)}, \ldots, \mu^{(r)})$. Put

$$\bar{X} = (\bar{X}^{(1)}, \ldots, \bar{X}^{(r)}) = n^{-1}\sum_{j=1}^{n} X_j,$$

$$S^{(s_1,s_2)} = n^{-1}\sum_{j=1}^{n}(X_j^{(s_1)} - \bar{X}^{(s_1)})(X_j^{(s_2)} - \bar{X}^{(s_2)})$$

and $S = (S^{(s_1,s_2)})$, an $r \times r$ matrix. Let $H$ be a real-valued function of $r$ variables, having a continuous derivative in a neighbourhood of $\mu$, and put $H_k(x) = (\partial/\partial x^{(k)})H(x)$, $\dot{H}(\bar{X}) = (H_1(\bar{X}), \ldots, H_r(\bar{X}))$ and $\hat{\sigma}^2 = \dot{H}(\bar{X})S\dot{H}(\bar{X})^T$, where the superscript $T$ denotes transpose. Now, $\hat{\theta} = H(\bar{X})$ is an estimator of $\theta = H(\mu)$, and $\hat{\sigma}$ is an estimator of the scale of $\hat{\theta}$. We consider approximations to the distribution of $(\hat{\theta} - \theta)/\hat{\sigma}$, which is an analogue of $(\bar{X} - \mu)/S$ (in the notation of Theorem 2.1).

Define $X_j^o = \sum_k X_j^{(k)} H_k(\mu)$, $\bar{X}^o = n^{-1}\sum_j X_j^o$ and $\mu^o = E(X_j^o)$, and let $X^{(k)}$ and $X^o$ have the distributions of $X_j^{(k)}$ and $X_j^o$, respectively. Put $\tau^{(k)}(x) = E\{X^{(k)^2} I(|X^{(k)}| \le x)\}$ and $\tau^o(x) = E\{X^{o^2} I(|X^o| \le x)\}$. We assume that the distributions of $X^{(1)}, \ldots, X^{(r)}$ and $X^o$ each satisfy conditions (2.1) and (2.2). Then there exist positive constants $b_n^{(k)}$ and $b_n^o$ such that $nb_n^{(k)^{-2}} \tau^{(k)}(b_n^{(k)}) \to 1$ and $nb_n^{o^{-2}} \tau^o(b_n^o) \to 1$. (See Proposition 5.1 for more detail of the properties of such sequences $b_n$.) It may be proved that $b_n^o = O(\max_{1\le k\le r} b_n^{(k)})$. We shall assume in addition that

$$\max_k b_n^{(k)} = O(b_n^o), \tag{4.5}$$

which is in effect equivalent to asking that $|\bar{X}^o - \mu^o|$ be the same size as $\max_k(|\bar{X}^{(k)} - \mu^{(k)}|)$ in probability. Under these conditions, we claim that the conclusions of Theorem 2.1 apply to $T_n = n^{1/2}(\hat{\theta} - \theta)/\hat{\sigma}$, and to its bootstrap version $T_m^*$.

Derivation of this result may proceed as follows. By Taylor expansion,

$$\hat{\theta} - \theta = H(\bar{X}) - H(\mu) = \bar{X}^o - \mu^o + o_p\left(\sum_{k=1}^{r} |\bar{X}^{(k)} - \mu^{(k)}|\right),$$

and $\hat{\sigma}^2 = S^{o^2} + o_p(\sum_k |\bar{X}^{(k)} - \mu^{(k)}|^2)$, where $S^{o^2}$ is the sample variance of $\{X_1^o, \ldots, X_n^o\}$. Condition (4.5) implies that $(\hat{\theta} - \theta)/\hat{\sigma} = (\bar{X}^o - \mu^o)/S^o + o_p(n^{-1/2})$. Since we have assumed that (2.1) and (2.2) hold for $X^o$, then the results of Theorem 2.1 apply to $T_n = n^{1/2}(\bar{X}^o - \mu^o)/S^o$, and hence also to $T_n = n^{-1/2}(\hat{\theta} - \theta)/\hat{\sigma}$. Similarly they apply to the bootstrap versions of these statistics.

## 5. Proof of Theorem 2.1

We begin by stating a proposition that holds under conditions a little more general than those of the theorem, but whose implications are not quite so transparent. For each positive $x$ put $\tau(x) = \tau_+(x) = E\{X^2 I(|X| \le x)\}$, and for integers $r \ge 2$ and values of $x$ such that $\tau(x) > 0$, define $\bar{X} = n^{-1}\sum X_j$, $S^2 = n^{-1}\sum(X_j - \bar{X})^2$, $T_n = n^{1/2}\bar{X}/S$ and

$$\rho_r(x, \lambda) = x^{2-r}\tau(x)^{-1} E\{X^r I(|X| \le \lambda x)\}.$$

Let $m = m(n)$ denote a sequence of positive integers diverging to infinity. If the mean of $X$ is finite then condition (5.3) below implies that the mean is zero, which explains why in our definition of $T_n$ we have not needed to centre the sample mean, $\bar{X}$. However, the conditions of the proposition do not demand finite mean; for example, they hold for any symmetric distribution in the domain of attraction of any stable law, without regard for the size of its exponent.

PROPOSITION 5.1. *Assume that there exists a sequence of positive constants $b_n$, diverging to infinity, such that the following conditions hold:*

$$nb_n^{-2}\tau(\lambda b_n) = O(1) \quad \text{for all} \quad \lambda > 0, \tag{5.1}$$

$$nb_n^{-2}\tau(b_n) \to 1, \tag{5.2}$$

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} nb_n^{-1}|E\{XI(|X| \le \lambda b_n)\}| = 0. \tag{5.3}$$

*Suppose too that for some $\delta > 0$,*

$$\lim_{\lambda\to\infty} \limsup_{x\to\infty} x^2\tau(x)^{-1}P(|X| > \lambda x) = 0, \tag{5.4}$$

$$\lim_{\epsilon\to 0} \epsilon^{\delta-1} \liminf_{x\to\infty} \tau(\epsilon x)\tau(x)^{-1} = \infty. \tag{5.5}$$

*Then*

$$\lim_{\epsilon\to 0,\lambda\to\infty} \liminf_{n\to\infty} P(\epsilon \le |T_n| \le \lambda) = 1, \tag{5.6}$$

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \sup_{-\infty<x<\infty} P(x < T_n \le x+\epsilon) = 0. \tag{5.7}$$

*If in addition to* (5.1)–(5.5),

$$|\rho_r(b_n,\lambda) - \rho_r(b_m,\lambda)| \to 0 \tag{5.8}$$

*for all sufficiently large $\lambda > 0$ and each integer $r \ge 2$, then*

$$\sup_{-\infty<x<\infty} |P(T_n \le x) - P(T_m \le x)| \to 0 \tag{5.9}$$

*as $n \to \infty$.*

PROOF. Define $U_n = b_n^{-1}\sum_{j=1}^n X_j$, $V_n = b_n^{-2}\sum_{j=1}^n X_j^2$. We shall prove that (5.1)–(5.5) imply that

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} P(|U_n| > \lambda) = 0, \tag{5.10}$$

$$\lim_{\epsilon\to 0,\lambda\to\infty} \liminf_{n\to\infty} P(\epsilon \le V_n \le \lambda) = 1, \tag{5.11}$$

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \sup_{-\infty<x<\infty} P(x < U_n \le x+\epsilon) = 0, \tag{5.12}$$

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \sup_{-\infty<x<\infty} P(x < V_n \le x+\epsilon) = 0, \tag{5.13}$$

and that if in addition (5.8) holds then

$$\sup_{-\infty<s,t<\infty} |P(U_n \le s, V_n \le t) - P(U_m \le s, V_m \le t)| \to 0. \tag{5.14}$$

Together these results imply the proposition in the case $\mu = 0$. To appreciate why, observe first that with

$$S_n = \left(\sum_{j=1}^{n} X_j\right) \Big/ \left(\sum_{j=1}^{n} X_j^2\right)^{1/2}$$

we have $T_n = S_n(1 - n^{-1}S_n^2)^{-1/2}$, and so it suffices to use (5.10)–(5.14) to establish that version of Proposition 5.1 in which $S_n$ replaces $T_n$ throughout. Let us interpret (5.6), (5.7) and (5.9) in that context. Then (5.6) and (5.7) follow directly from (5.10)–(5.12), and (5.9) from (5.10)–(5.12) and (5.14). (We shall use both (5.12) and (5.13) during the proof of (5.14).) For example, to derive (5.9), first use (5.10) and (5.11) to show that it is sufficient to prove that for each $\epsilon > 0$ and $x_0 > 0$, and each integer $p \geq 1$,

$$\sup_{-x_0 < x < x_0} |P(S_n \leq x, \epsilon < V_n \leq p\epsilon) - P(S_m \leq x, \epsilon < V_m \leq p\epsilon)| \to 0. \tag{5.15}$$

To derive (5.15), suppose we have already proved (5.7), and let $\delta > 0$ (different from the $\delta$ in (5.5)) denote a divisor of $\epsilon$, so that $j_0 = 1 + (\epsilon/\delta)$ is an integer. Put $\delta' = 2x_0/j_0$, which quantity tends to zero as $\delta$ decreases to zero. Then if $x > 0$,

$$\begin{aligned}
&P(S_n \leq x, \epsilon < V_n \leq p\epsilon) \\
&\quad = \sum_{j=j_0}^{p(j_0-1)} P\{U_n/V_n \leq x, V_n \in ((j-1)\delta, j\delta]\} \\
&\quad \leq \sum_{j=j_0}^{p(j_0-1)} P\{U_n \leq j\delta x, V_n \in ((j-1)\delta, j\delta]\} \\
&\quad = \sum_{j=j_0}^{p(j_0-1)} P\{U_m \leq j\delta x, V_m \in ((j-1)\delta, j\delta]\} + o(1) \\
&\quad \leq \sum_{j=j_0}^{p(j_0-1)} P\{U_m/V_m \leq j(j-1)^{-1}x, V_m \in ((j-1)\delta, j\delta]\} + o(1) \\
&\quad \leq \sum_{j=j_0}^{p(j_0-1)} P\{U_m/V_m \leq x, V_m \in ((j-1)\delta, j\delta]\} \\
&\qquad + \sup_{j_0 \leq j \leq p(j_0-1)} P\{x < U_m/V_m \leq j(j-1)^{-1}x\} + o(1) \\
&\quad \leq P(S_m \leq x, \epsilon < V_m < p\epsilon) + \sup_{-\infty < x < \infty} P(x < S_m \leq x + \delta') + o(1),
\end{aligned}$$

where the remainder term is of the stated order uniformly in $0 < x \leq x_0$, for each fixed $\delta > 0$. (The third relation is a consequence of (5.14), and the others follow by elementary arguments.) In view of (5.7) the last-written supremum may be made arbitrarily small by choosing $\delta$ sufficiently small, and the case $x > 0$ may be treated similarly. Arguing in this way we may show that

$$\limsup_{n\to\infty} \sup_{|x|\le x_0} \{P(S_n \le x, \epsilon < V_n \le p\epsilon) - P(S_m \le x, \epsilon < V_m \le p\epsilon)\} = 0.$$

The counterpart of this result, in which "lim sup" and "sup" are replaced by "lim inf" and "inf" respectively, may be proved similarly. Together they imply (5.15).

We now proceed to derive (5.10)–(5.13). Observe that if $\eta > 0$ is given then by (5.4) we may choose $\lambda$ so large that $P(|X| > \lambda b_n) < \eta b_n^{-2}\tau(b_n)$ for all sufficiently large $n$. It follows that

$$P(|U_n| > \lambda) \le nP(|X| > \lambda b_n) \sim b_n^2 \tau(b_n)^{-1} P(|X| > \lambda b_n) < \eta,$$

the second relation following from (5.2). This proves (5.10). Similarly,

$$P(V_n > \lambda) \le nP(|X| > \lambda b_n) \le \eta + o(1),$$

whence (5.11) will follow if we prove that

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} P(V_n \le \epsilon^2) = 0. \tag{5.16}$$

To this end, put $Y_j = X_j^2 I(|X_j| \le \epsilon b_n) - \tau(\epsilon b_n)$ and $x_n = n\tau(\epsilon b_n) - b_n^2\epsilon^2$, and note that by (5.5), $x_n > 0$ for all sufficiently large $n$, provided $\epsilon < 1$ is chosen sufficiently small. Now,

$$\begin{aligned} P(V_n \le \epsilon^2) &\le P\left(\sum_{j=1}^{n} Y_j \le -x_n\right) \le n x_n^{-2} E(Y_1^2) \\ &\le n x_n^{-2} b_n^2 \epsilon^2 \tau(b_n) \sim x_n^{-2} b_n^4 \epsilon^2 \sim \epsilon^2\{\tau(b_n)^{-1}\tau(\epsilon b_n) - \epsilon^2\}^{-2}. \end{aligned}$$

Assumption (5.5) implies that for each $\eta > 0$, $\epsilon^2\{\tau(b_n)^{-1}\tau(\epsilon b_n) - \epsilon^2\}^{-2}$ may be made less than $\eta$ for all large $n$ by choosing $\epsilon$ sufficiently small. This proves (5.16).

Next we establish (5.12). For that purpose we employ a bound for the concentration function:

$$\sup_{-\infty<x<\infty} P(x \le U_n \le x + \epsilon) \le 2(96/95)^2 \epsilon \int_0^{1/\epsilon} |E \exp(iub_n^{-1}X)|^n du. \tag{5.17}$$

See Petrov ((1975), p. 38). Put $X^s = X_1 - X_2$ (the symmetrized version of $X$). Since $1 - \cos x \ge (1/4)x^2$ for any real $x$ satisfying $|x| \le 1$ then for any real $t$, and with $v = (2t)^{-1}$,

$$\begin{aligned} |E \exp(itX)|^2 &= 1 - E\{1 - \cos(tX^s)\} \\ &\le 1 - (1/4)E\{(tX^s)^2 I(|tX^s| \le 1)\} \\ &\le 1 - (1/2)t^2[E\{X^2 I(|X| \le v)\}P(|X| \le v) \\ &\qquad - \{EXI(|X| \le v)\}^2]. \end{aligned}$$

Hence,

$$|E\exp(itX)|^n \le \exp(-(1/4)nt^2\tau(v) \\ + nt^2[\tau(v)P(|X|>v) + \{EXI(|X|\le v)\}^2]).$$

Replacing $t$ by $u/b_n$ and noting (5.1)–(5.3) we deduce that

$$|E\exp(iub_n^{-1}X)|^n \le \exp\{-(1/4)u^2\tau(b_n)^{-1}\tau(b_n/2u) + o(1)\},$$

where the $o(1)$ term is of that order uniformly in $u_1 \le u \le u_2$, for any $0 < u_1 < u_2 < \infty$. Therefore, in view of (5.5),

$$\limsup_{n\to\infty} \int_{u_1}^{u_2} |E\exp(iub_n^{-1}X)|^n du \le C_1 + \int_{u_1}^{u_2} \exp(-C_2 u^{1+\delta})du \le C_3$$

uniformly in $0 < u_1 < u_2 < \infty$, where $C_1$, $C_2$ and $C_3$ are positive constants not depending on $u_1$ or $u_2$, and $\delta > 0$ is as in (5.5). (We have applied (5.5) with $x = b_n$ and $\epsilon = (2u)^{-1}$.) Hence, the lim sup as $n \to \infty$ of the integral of $|E\exp(iub_n^{-1}X)|^n$ over $u$ in any interval $(u_1, u_2)$ with $0 < u_1 < u_2 < \infty$, is bounded uniformly in $u_1$ and $u_2$. Result (5.12) now follows from (5.17), and (5.13) may be proved similarly.

As a prelude to deriving (5.14) we establish an approximation to the joint characteristic function of $(U_n, V_n)$,

$$\psi_n(u,v) = E\{\exp(iuU_n + ivV_n)\} = \chi(u,v)^n,$$

where $\chi(u,v) = E\{\exp(iub_n^{-1}X + ivb_n^{-2}X^2)\}$ and $i$ denotes the square root of $-1$. Given an integer $k \ge 1$ and any $\lambda > 0$, put

$$\delta_{nk}(u,v) = \sum_{j=1}^{k} \frac{(-1)^j}{(2j)!} E\{(ub_n^{-1}X + vb_n^{-2}X^2)^{2j} I(|X| \le \lambda b_n)\} \\ + i\sum_{j=0}^{k} \frac{(-1)^j}{(2j+1)!} E\{(ub_n^{-1}X + vb_n^{-2}X^2)^{2j+1} I(|X| \le \lambda b_n)\}.$$

We shall prove that for each $u_0, v_0, \eta > 0$ there exist $\lambda, k > 0$, chosen sufficiently large, such that for all $n$, $nb_n^{-1}|E\{XI(|X| \le \lambda b_n)\}| \le \eta/u_0$ and

$$\sup_{|u|\le u_0, |v|\le v_0} |\psi_n(u,v) - \exp\{n\delta_{nk}(u,v)\}| \le \eta. \tag{5.18}$$

The first step is to note that, in view of (5.2)–(5.4), for any $\xi > 0$ we may choose $\lambda$ so large that $nP(|X| > \lambda b_n) \le \xi/2$ and $nb_n^{-1}|E\{XI(|X| \le \lambda b_n)\}| \le \eta/u_0$ for all $n$. Fix this $\lambda$, and let $\Lambda$ be an upper bound to

$$nE\{(ub_n^{-1}X + vb_n^{-2}X^2)^2 I(|X| \le \lambda b_n)\}$$

uniformly in $|u| \le u_0$ and $|v| \le v_0$. That there exists such a $\Lambda$, not depending on $n$, follows from (5.1) and (5.2). Let $\mathcal{E}$ denote the event that $|X| \le \lambda b_n$, with $\tilde{\mathcal{E}}$ being the complementary event. Put $Z_n = iub_n^{-1}X + ivb_n^{-2}X^2$. Then $|E\{\exp(Z_n)I(\tilde{\mathcal{E}})\}| \le \xi/(2n)$, and on $\mathcal{E}$, $|Z_n| \le \zeta \equiv u_0\lambda + v_0\lambda^2$, uniformly in $|u| \le u_0$ and $|v| \le v_0$. It follows that

$$\left|\exp(Z_n) - \sum_{j=0}^{2k+1}(j!)^{-1}Z_n^j\right| \le \{(2k+2)!\}^{-1}\zeta^{2k}e^{\zeta}|Z_n|^2$$

on $\mathcal{E}$. Choose $k$ so large that $\{(2k+2)!\}^{-1}\zeta^{2k}e^{\zeta}\Lambda \le \xi/2$. Then

$$E\left\{\left|\exp(Z_n) - \sum_{j=0}^{2k+1}(j!)^{-1}Z_n^j\right| I(\mathcal{E})\right\} \le \xi/(2n).$$

Therefore,

$$\left|E\left\{\exp(Z_n) - \sum_{j=0}^{2k+1}(j!)^{-1}Z_n^j I(\mathcal{E})\right\}\right| \le \xi/n. \tag{5.19}$$

A similar argument may be employed to show that for a constant $C > 0$, depending on $u_0$, $v_0$, $\lambda$ and $k$,

$$\left|E\left\{\sum_{j=1}^{2k+1}(j!)^{-1}Z_n^j I(\mathcal{E})\right\}\right| \le C/n. \tag{5.20}$$

Together the bounds (5.19) and (5.20) may be used to prove that for all sufficiently large $n$, and for each $u_0, v_0 > 0$,

$$\sup_{|u|\le u_0, |v|\le v_0}\left|\{E\exp(Z_n)\}^n - E\exp\left\{n\sum_{j=1}^{2k+1}(j!)^{-1}Z_n^j I(\mathcal{E})\right\}\right| \le \eta,$$

provided $\xi$ is chosen sufficiently small. This is equivalent to (5.18).

Consider a term-by-term expansion of $n\delta_{nk}(u,v)$, producing a polynomial in $(u,v)$ with coefficients given by constant multiples of quantities of the form

$$\nu_{nr} = nb_n^{-r}E\{X^r I(|X| \le \lambda b_n)\} = \{1+o(1)\}\rho_r(b_n,\lambda),$$

where $1 \le r \le 2(2k+1)$. If $r \ge 2$ then $|\rho_r(b_n,\lambda)| \le \lambda^{r-2}\tau(\lambda b_n)/\tau(b_n) = O(1)$, by (5.1) and (5.2), and so by (5.8), $\nu_{nr} - \nu_{mr} \to 0$. When $r = 1$ we have, by choice of $\lambda$, that $|\nu_{n1} - \nu_{m1}| \le |\nu_{n1}| + |\nu_{m1}| \le 2\eta$. Therefore, given $\eta' > 0$, $|\exp\{n\delta_{nk}(u,v)\} - \exp\{m\delta_{mk}(u,v)\}|$ may be made less than $\eta'$ uniformly in $|u| \le u_0$ and $|v| \le v_0$, by choosing $\eta$ sufficiently small and then $n$ sufficiently large. Hence, (5.18) implies that

$$\sup_{|u|\le u_0, |v|\le v_0}|\psi_n(u,v) - \psi_m(u,v)| \to 0 \tag{5.21}$$

as $n \to \infty$.

To complete the proof of (5.14), let $\epsilon > 0$ and write $M$ and $N$ for independent Normal $N(0, \epsilon)$ random variables, independent also of $U_n$ and $V_n$. Put $U_n^\dagger = U_n + M$ and $V_n^\dagger = V_n + N$, and note that the characteristic function of $(U_n^\dagger, V_n^\dagger)$ is given by $\psi_n^\dagger(u,v) = \psi_n(u,v)\exp\{-(1/2)\epsilon(u^2+v^2)\}$. In view of (5.21) and the fact that $|\psi_n| \le 1$,

$$\sup_{|u| \le u_0, |v| \le v_0} |\psi_n^\dagger(u,v) - \psi_m^\dagger(u,v)| \to 0$$

as $n \to \infty$. It now follows by Fourier inversion of the characteristic functions $\psi_n^\dagger$ and $\psi_m^\dagger$ that for each fixed $\epsilon$,

$$\sup_{\mathcal{C}} |P\{(U_n^\dagger, V_n^\dagger) \in \mathcal{C}\} - P\{(U_m^\dagger, V_m^\dagger) \in \mathcal{C}\}| \to 0, \tag{5.22}$$

where the supremum is taken over all convex sets $\mathcal{C} \subseteq \mathbb{R}^2$.

We claim that

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \sup_{-\infty < s,t < \infty} |P(U_n \le s, V_n \le t) - P(U_n^\dagger \le s, V_n^\dagger \le t)| = 0. \tag{5.23}$$

Together, (5.22) and (5.23) imply (5.14). To prove (5.23) observe that for any $\delta > 0$ and all $-\infty < s,t < \infty$,

$$\begin{aligned} P(U_n^\dagger \le s, V_n^\dagger \le t) &\le P(U_n \le s+\delta, V_n \le t+\delta) + 2P(|N| > \delta) \\ &\le P(U_n \le s, V_n \le t) + 2P(|N| > \delta) \\ &\quad + \sup_{-\infty < x < \infty} \{P(x < U_n \le x+\delta) + P(x < V_n \le x+\delta)\}. \end{aligned}$$

Given $\eta > 0$, choose $\delta$ so small that for all sufficiently large $n$, the supremum on the right-hand side does not exceed $\eta/3$. (For this purpose, (5.12) and (5.13) are required.) Next select $\epsilon_0$ so small that for all $\epsilon < \epsilon_0$, $P(|N| > \delta) < \eta/3$. It follows that if $\epsilon < \epsilon_0$ then for all sufficiently large $n$,

$$\sup_{-\infty < s,t < \infty} \{P(U_n^\dagger \le s, V_n^\dagger \le t) - P(U_n \le s, V_n \le t)\} \le \eta.$$

The counterpart of this formula, in which "sup" is replaced by "inf" and the relation "$\le \eta$" is replaced by "$\ge -\eta$", may be derived by a similar argument. Result (5.23) is an immediate consequence. This concludes the proof of Proposition 5.1.

Our next result is in a sense an empirical version of Proposition 5.1.

PROPOSITION 5.2. *Assume the conditions of Proposition* 5.1, *except that we replace* (5.8) *by either*

$$mn^{-1} + m^{-1} \to 0 \tag{5.24}$$

*(in the case of a weak law) or by the slightly more stringent assertion*

$$mn^{-1}\log n + m^{-1} \to 0 \tag{5.25}$$

*(for a strong law). If* (5.24) *holds then*

$$\sup_{-\infty<x<\infty} |P(T_m^* \le x|\mathcal{X}) - P(T_m \le x)| \to 0 \tag{5.26}$$

*in probability, and if in addition* (5.25) *is true then* (5.26) *holds with probability one.*

PROOF. Put $U_m^* = b_m^{-1}\sum_{j=1}^m (X_j^* - \bar{X})$, $V_m^* = b_m^{-2}\sum_{j=1}^m (X_j^* - \bar{X})^2$ and

$$S_m^* = \left\{\sum_{j=1}^m (X_j^* - \bar{X})\right\} \Big/ \left\{\sum_{j=1}^n (X_j^* - \bar{X})^2\right\}^{1/2}.$$

(To simplify matters we have dropped the extra subscript $n$ from the ideal notation $U_{nm}^*$ and $V_{nm}^*$.) Using a modified form of the argument outlined at the beginning of the proof of Proposition 5.1 we may show that it suffices to establish that version of Proposition 5.2 in which $T_m^*$ is replaced by $S_m^*$, and that for this it is enough to prove that

$$\sup_{-\infty<s,t<\infty} |P(U_m^* \le s, V_m^* \le t|\mathcal{X}) - P(U_m \le s, V_m \le t)| \to 0, \tag{5.27}$$

where the convergence is in probability under (5.24) and with probability one if (5.25) holds. Note that results (5.10)–(5.13) derived during the proof of Proposition 5.1 remain valid under the assumptions of Proposition 5.2.

Let $\hat{\psi}_m$ denote the conditional characteristic function of $(U_m^*, V_m^*)$. We shall prove that for each $u_0, v_0 > 0$,

$$\int_{|u|\le u_0, |v|\le v_0} |\hat{\psi}_m(u,v) - \psi(u,v)|\,du\,dv \tag{5.28}$$

converges to zero as $n \to \infty$, the mode of convergence being in probability if (5.24) holds and with probability one under (5.25). Result (5.27) follows from (5.28) on using a modified form of the smoothing argument employed in the last two paragraphs of the proof of Proposition 5.1, as follows. First, observe that in view of (5.17) and its analogues for $U_n^*$, $V_n$ and $V_n^*$, the bootstrap versions of (5.12) and (5.13) follow from (5.28) via the argument in the paragraph containing (5.17). Therefore,

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\sup_{-\infty<x<\infty} P(x < U_m^* \le x + \epsilon|\mathcal{X}) = 0 \tag{5.29}$$

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\sup_{-\infty<x<\infty} P(x < V_m^* \le x + \epsilon|\mathcal{X}) = 0. \tag{5.30}$$

Next, let $M$ and $N$ be as they were in the proof of Proposition 5.1, except that they must now be conditionally independent of the $X_j^*$'s as well as independent of the $X_j$'s. Put $U_n^\dagger = U_n + M$, $V_n^\dagger = V_n + N$, $U_n^\ddagger = U_n^* + M$ and $V_n^\ddagger = V_n^* + N$. The analogue of (5.22),

$$(5.31) \qquad \sup_{C} |P\{(U_m^\ddagger, V_m^\ddagger) \in \mathcal{C}|\mathcal{X}\} - P\{(U_m^\dagger, V_m^\dagger) \in \mathcal{C}\}| \to 0,$$

follows from (5.28) by Fourier inversion. Formula (5.27) follows from (5.31), (5.23) and the bootstrap version of the latter, i.e.

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \sup_{-\infty<s,t<\infty} |P(U_m^* \le s, V_m^* \le t|\mathcal{X}) - P(U_m^\ddagger \le s, V_m^\ddagger \le t|\mathcal{X})| = 0.$$

That result may be derived in the manner outlined in the last paragraph of the proof of Proposition 5.1, using (5.29) and (5.30) in place of (5.13) and (5.14).

It remains only to prove (5.28). Let $\mathcal{X}$, the characteristic function of $(b_n^{-1}X, b_n^{-2}X^2)$, be as in the proof of Proposition 5.1, and let $\hat{\chi}$ denote its bootstrap estimator,

$$\hat{\chi}(u,v) = n^{-1}\sum_{j=1}^{n} \exp(iub_n^{-1}X_j + ivb_n^{-2}X_j^2).$$

Then $\psi_m = \chi^m$ and $\hat{\psi}_m = \hat{\chi}^m$, so that

$$|\hat{\psi}_m - \psi_m| = |(\hat{\chi} - \chi)(\hat{\chi}^{m-1} + \hat{\chi}^{m-2}\chi + \cdots + \chi^{m-1})| \le m|\hat{\chi} - \chi|.$$

An alternative bound, $|\hat{\chi} - \chi| \le 2$, is obvious. Therefore, writing $\mathcal{R}$ for the rectangle in $(u,v)$ space defined by $|u| \le u_0$ and $|v| \le v_0$, we have for each $\eta > 0$,

$$\begin{aligned}
\int_{\mathcal{R}} |\hat{\psi}_m - \psi_m| &\le \int_{\mathcal{R}} \min(m|\hat{\chi} - \chi|, 2) \\
&\le 8\eta u_0 v_0 + 2\int_{\mathcal{R}} I(|\hat{\chi} - \chi| > 2\eta/m) \\
&\le 8\eta u_0 v_0 + 2\int_{\mathcal{R}} I\{|\Re(\hat{\chi} - \chi)| > \eta/m\} \\
&\quad + 2\int_{\mathcal{R}} I\{|\Im(\hat{\chi} - \chi)| > \eta/m\}.
\end{aligned}$$

Therefore it suffices to prove that for each $\eta > 0$,

$$(5.32) \qquad \begin{aligned}
&\int_{\mathcal{R}} I\{|\Re(\hat{\chi} - \chi)| > \eta/m\} \to 0, \\
&\int_{\mathcal{R}} I\{|\Im(\hat{\chi} - \chi)| > \eta/m\} \to 0.
\end{aligned}$$

We shall derive only (5.32). Indeed, with $A_j = ub_n^{-1}X_j + vb_n^{-2}X_j^2$ and $\tilde{\chi}(u,v) = n^{-1}\sum_{j=1}^{n} \exp(iA_j)$ we shall prove that

$$(5.33) \qquad \int_{\mathcal{R}} I\{|\Re(\tilde{\chi} - \chi)| > \eta/m\} \to 0,$$

there being a similar proof that

$$\int_{\mathcal{R}} I\{|\Re(\tilde{\chi} - \hat{\chi})| > \eta/m\} \to 0.$$

Put $B_j = 1 - \cos A_j$, and observe that

$$|\Re(\hat{\chi} - \chi)| = n^{-1} \left| \sum_{j=1}^{n} (B_j - EB_j) \right|.$$

The variables $B_j$ are independent and identically distributed with $|B_j| \leq 2$, and for any $\lambda > 0$,

$$E(B_j^2) \leq E\{A_j^2 I(|X_j| \leq \lambda b_n)\} + 4P(|X_j| > \lambda b_n).$$

It follows from this inequality, (5.1), (5.2) and (5.4) that if $\lambda > 0$ is sufficiently large then $E(B_j^2) \leq C_1 n^{-1}$, uniformly in $|u| \leq u_0$ and $|v| \leq v_0$, where $C_k$ denotes a constant depending only on $\lambda$, $u_0$, $v_0$ and, in the work below, $\eta$. Therefore, by Hoeffding's inequality (e.g. Pollard (1984), pp. 191–192),

$$P[|\Re\{\hat{\chi}(u,v) - \chi(u,v)\}| > x/n] \leq 2\exp\{-C_2 x^2 (1+x)^{-1}\}$$

uniformly in $|u| \leq u_0$, $|v| \leq v_0$ and $x > 0$. Hence,

$$E\left[\int_{\mathcal{R}} I\{|\Re(\hat{\chi} - \chi)| > \eta/m\}\right] \leq C_3 \exp(-C_4 n/m).$$

This implies (5.33). (Use the Borel-Cantelli lemma to obtain strong convergence in (5.33) when (5.25) holds.) The proof of Proposition 5.2 is complete.

Combining Propositions 5.1 and 5.2 we see that under conditions (5.1)–(5.5), results (2.4) and (2.5) hold; that under (5.1)–(5.5) and either (5.24) or (5.25) (the latter if we require strong convergence), (2.3) is valid; and that if we assume in addition (5.8) then (2.6) holds. Therefore, in order to establish Theorem 2.1 it suffices to prove that the conditions in Theorem 2.1 imply those in Propositions 5.1 and 5.2.

PROPOSITION 5.3. *Condition* (2.1) *implies* (5.1)–(5.5), *and* (2.1) *and* (2.2) *imply the existence of a sequence of positive constants* $\delta_n$, *decreasing to zero and such that* (5.8) *holds whenever* $\delta_n \leq m/n \to 0$.

PROOF. Without loss of generality, $\mu = E(X) = 0$. Recall that we write $\tau$ for $\tau_+$. To obtain (5.2), observe that by the first part of (2.1), if $\eta > 0$ is given then we may choose $\lambda > 1$ so close to 1 that for all sufficiently large $x$,

$$(x/\lambda)^2 P(|X| = x) \leq \tau(x) - \tau(x/\lambda) \leq \eta\tau(x).$$

Therefore, $x^2P(|X| = x)/\tau(x) \to 0$ as $x \to \infty$. It follows that there exists a sequence of constants $b_n$ such that (5.2) holds. Result (5.1) follows from (2.1) and (5.2).

To derive (5.3), observe that by the second part of (2.1) there exists a constant $C_1 > 0$ such that

$$xE\{|X|I(|X| > x)\} = \int_1^\infty u^{-2}\tau(ux)du \le C_1\tau(x)\int_1^\infty u^{-1-\epsilon}du.$$

Now replace $x$ by $\lambda x$, and again apply (2.1), to deduce that for another constant $C_2 > 0$, not depending on $\lambda$ or $x$, we have for all large $x$, $xE\{|X|I(|X| > \lambda x)\} \le C_2\lambda^{-\epsilon}\tau(x)$. Result (5.3) follows on substituting $b_n$ for $x$ in this inequality, and noting (5.2).

Result (5.5) is an immediate consequence of the second part of (2.1). To derive (5.4), observe that

$$x^2P(|X| > x) \le 2\int_1^\infty u^{-3}\tau(ux)du \le C_1\tau(x)\int_2^\infty u^{-2-\epsilon}du.$$

Replace $x$ by $\lambda x$ and apply (2.1) again, to obtain the result that $(\lambda x)^2P(|X| > \lambda x) \le C_2\lambda^{1-\epsilon}\tau(x)$, whence follows (5.4).

Next, observe that

$$\begin{aligned}\rho_r(b_n,\lambda) - \rho_r(b_m,\lambda) = (r-2)\int_0^1 u^{r-3}\{&\rho_\pm(b_n,\lambda) - \rho_\pm(b_n,\lambda u)\\ &- \rho_\pm(b_m,\lambda) + \rho_\pm(b_m,\lambda u)\}du,\end{aligned}$$

where the $\pm$ signs are chosen according to whether $r$ is even or odd, respectively. Therefore, (5.8) is true if

$$\rho_\pm(b_n,\lambda) - \rho_\pm(b_m,\lambda) \to 0 \tag{5.34}$$

for each $\lambda > 0$. In view of (2.2) we may choose $\zeta = \zeta(x)$ such that $\zeta(x) \to 0$ as $x \to \infty$, and (2.2) continues to hold (for each $u$) with this $\zeta$. Then (5.34) is valid if $m/n \to 0$ and $b_m/b_n \ge \zeta(b_n)$. The existence of a sequence $\{\delta_n\}$ such that (a) $n\delta_n \to \infty$ and (b) $m/n \ge \delta_n$ and $m/n \to 0$ imply (5.8) may now be proved from (5.1) and (5.2), using an argument based on reduction to contradiction.

This completes the proof of Proposition 5.3. Theorem 2.1 follows from Propositions 5.1–5.3.

## Acknowledgements

We are grateful for the helpful comments of two referees, which have led to significant improvements in presentation.

## REFERENCES

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

Arcones, M. A. and Giné, E. (1991). Additions and corrections to the bootstrap of the mean with arbitrary sample size, *Ann. Inst. Henri Poincaré Ser. B*, **27**, 583–595.

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

Deheuvels, P., Mason, D. M. and Shorack, G. R. (1993). Some results on the influence of extremes on the bootstrap, *Ann. Inst. Henri Poincaré Ser. B*, **29**, 83–103.

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

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

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

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

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

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, **75**, 237–249.

Owen, A. B. (1990). Empirical likelihood confidence regions, *Ann. Statist.*, **18**, 90–120.

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

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

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works, *Comm. Statist. Theory Methods*, **15**, 3193–3203.

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