Bernoulli **6**(5), 2000, 809–834

Bootstrap relative errors and sub-exponential distributions

ANDREW T. A. WOOD

School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, UK.
E-mail: Andy.Wood@nottingham.ac.uk

For the purposes of this paper, a distribution is sub-exponential if it has finite variance but its moment generating function is infinite on at least one side of the origin. The principal aim here is to study the relative error properties of the bootstrap approximation to the true distribution function of the sample mean in the important sub-exponential cases. Our results provide a fairly general description of how the bootstrap approximation breaks down in the tail when the underlying distribution is sub-exponential and satisfies some very mild additional conditions. The broad conclusion we draw is that the accuracy of the bootstrap approximation in the tail depends, in a rather sensitive way, on the precise tail behaviour of the underlying distribution. Our results are applied to several sub-exponential distributions, including the lognormal. The lognormal case is of particular interest because, as the simulation studies of Lee and Young have shown, bootstrap confidence intervals can have very poor coverage accuracy when applied to data from the lognormal.

Keywords: Edgeworth expansion; moderate deviations; percentile method; tail probability

1. Introduction

Since Efron's (1979) landmark paper, the theoretical properties of the bootstrap have been studied extensively using the technique of Edgeworth expansion; see Singh (1981), Hall (1986; 1988) and, especially, the monograph by Hall (1992). However, although this work has provided much valuable insight into the theoretical properties of the bootstrap, it is fair to say that the Edgeworth perspective does not give a complete theoretical picture. Hall (1992, p. 323) makes the following remark: 'One could be excused for thinking that the Edgeworth view [of the bootstrap] was the only one, or that it provided all the information that we need about properties of the bootstrap in problems of distribution estimation. This is certainly not the case.' A key feature of the Edgeworth view of the bootstrap is that it focuses on *absolute*, as opposed to *relative*, errors. Our belief is that a relative error view of the bootstrap is an important complement to an absolute error (i.e. Edgeworth) view. The purpose of this paper is to present new results on relative error properties of the bootstrap.

The study of relative error properties of the bootstrap when the relevant exponential moments are finite was initiated by Hall (1990; 1992, p. 324); see also Jing (1992), Jing *et al*. (1994) and Booth *et al*. (1994) for a selection of relative error results for the bootstrap. However, as far as the author is aware, there is nothing in the literature on relative error properties of the bootstrap in the *subexponential* case, i.e. when the population variance is

1350–7265

finite but the relevant moment generating function is infinite on at least one side of the origin.

An important example of a sub-exponential distribution is the lognormal. The extensive numerical results of Lee and Young (1995) show that the coverage accuracy of bootstrap confidence intervals when the data come from a lognormal distribution is sometimes very poor, even disastrous. An absolute error view of the bootstrap does not provide clear theoretical insight into why this should be so – though the calculation of coefficients in Edgeworth expansions, as performed in Lee and Young (1995), can provide useful diagnostic information. In constrast, the relative error results presented in this paper show that the bootstrap is on much less favourable ground when applied to data from a lognormal distribution, compared with distributions whose moment generating functions are finite on both sides of the origin.

A comment on terminology. Some authors use the term *subexponential* for a class $\mathscr{S}$ of distribution functions defined as follows: $F \in \mathscr{S}$ if and only if

$$F(0+) = 0; \qquad F(x) < 1 \quad \text{for all } x > 0; \qquad F(\infty) = 1;$$
$$1 - F^{(2)}(x) \sim 2\{1 - F(x)\} \quad \text{as } x \to \infty, \tag{1.1}$$

where

$$F^{(2)}(x) = \int_0^\infty F(x-y)\,\mathrm{d}F(y)$$

is the convolution of F with itself; see Teugels (1975), Pitman (1980) and Bingham *et al.* (1987). The class $\mathscr{S}$ is different from the class of sub-exponential distributions considered here (and note that the hyphen is used to distinguish our class from $\mathscr{S}$). In particular, (1.1) implies that F is the distribution function of a non-negative random variable, whereas we do not impose any non-negativity condition; and, on the other hand, $\mathscr{S}$ contains distributions with arbitrarily heavy tails, whereas we insist that some power moments are finite. Nevertheless, it does seem (though we have not checked this systematically) that the results in this paper cover most (and possibly all) of the 'interesting' distributions in $\mathscr{S}$ which have finite fourth moment.

A principal application of the bootstrap idea is to the construction of nonparametric confidence intervals for a population quantity of interest. Numerous variants of the bootstrap have been developed for this purpose, including: ordinary percentile intervals, percentile-t intervals, ABC intervals and intervals based on iterative forms of the bootstrap; see Hall (1992) and Efron and Tibshirani (1993) for further details.

In this paper we shall focus our attention on the simpler question of bootstrap approximation to the true distribution of a sample mean. One may ask how the relative error results that we present for this case extend to more complicated statistics (such as a studentized sample mean), and translate into corresponding results for the relative coverage error of various types of bootstrap confidence interval. These are open questions which are currently under investigation.

Let $\mathscr{X} = \{X_1, \ldots, X_n\}$ denote a random sample of size n from an underlying population with distribution function F, mean μ and variance σ^2. Let $\mathscr{X}^* =$

$\{X_1^*, \ldots, X_n^*\}$ denote a (re)sample obtained by sampling randomly with replacement from $\mathscr{X}$. Write $\overline{X}$ and $\overline{X}^*$ for the mean of the samples $\mathscr{X}$ and $\mathscr{X}^*$, respectively. In this paper we study the quality of the bootstrap approximation $P[n^{1/2}(\overline{X}^* - \overline{X}) > y|\mathscr{X}]$ to the 'true' tail probability $P[n^{1/2}(\overline{X} - \mu) > y]$. Note that the former probability can be estimated with arbitrary accuracy by resampling from $\mathscr{X}$, whereas the latter probability is unknown if F is unknown.

The *absolute error* of this bootstrap approximation is given by

$$|P[n^{1/2}(\overline{X}^* - \overline{X}) > y|\mathscr{X}] - P[n^{1/2}(\overline{X} - \mu) > y]|,$$

typically for fixed y. In a *relative error* approach, one studies the ratio

$$P[n^{1/2}(\overline{X}^* - \overline{X}) > y_n|\mathscr{X}]/P[n^{1/2}(\overline{X} - \mu) > y_n] \tag{1.2}$$

where $y_n \to \infty$ at a suitable rate as $n \to \infty$. Although from a practical point of view it may seem artificial to allow y_n to depend on n and increase to infinity, it turns out that this approach does produce interesting information about the behaviour of the bootstrap approximation in the tail. This information does not come to light when y is held fixed.

The principal purpose of this paper is to study (1.2) when the distribution of X_1 is sub-exponential. The particular question we focus on is: what is the critical rate of increase of $y = \lambda_n$ at which the bootstrap approximation breaks down? If the distribution of X_1 has a moment generating function which is finite in a neighbourhood of the origin, the result is already known: see Hall (1990; 1992, Appendix V) and Theorem 2.2 below. However, as far as we are aware, nothing has been published on the case in which X_1 has a sub-exponential distribution.

We now briefly outline our main results. In Theorem 2.1, we present a general result on the behaviour of the numerator in (1.2); in Theorem 2.2, we recall Hall's result in a slightly different form; and in Theorems 2.3–2.5, we describe the behaviour of the denominator in (1.2) in the three most important sub-exponential cases. In Corollaries 2.1–2.3, we give a simple characterization of the critical rate λ_n at which the bootstrap approximation breaks down in these three cases.

Theorems 2.3–2.5 are concerned with moderate and large deviation probabilities for the sample mean when the underlying population is sub-exponential. There is a substantial and impressive Russian literature on this topic; see S.V. Nagaev (1979) for a review. However, apart from Theorem 2.3, which is due to A.V. Nagaev (1969), the results given by Nagaev (1979) are not presented in a form which is suitable for our purposes. For this reason, we have formulated and proved new results, Theorems 2.4 and 2.5, which directly address our needs. The main ingredients of the proofs of these two theorems are condition (2.5), Lemma 4.2 and Proposition 4.2.

In Section 3, we study several sub-exponential examples, including the lognormal. In Section 4, we establish several auxiliary results, and Theorems 2.1, 2.4 and 2.5 are proved in Section 5. A small simulation study is described in Section 6, and we conclude with a brief discussion in Section 7.

2. Main results

The following notation will be used throughout the paper. Consider a sample $\mathscr{X} = \{X_1, \ldots, X_n\}$ of independent, identically distributed random variables with corresponding order statistics $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$. Define

$$S_n = \sum_{i=1}^{n}(X_i - \mu) \quad \text{and} \quad G_n(y) = P[n^{-1/2}S_n > y],$$

where $\mu = \mathrm{E}(X_1)$ will usually be assumed to be zero. The corresponding bootstrap quantities are defined analogously: let $X_1^*, \ldots, X_n^*$ be a resample drawn randomly, with replacement, from $\mathscr{X}$ and define

$$S_n^* = \sum_{i=1}^{n}(X_i^* - \overline{X}) \quad \text{and} \quad \hat{G}_n(y) = P[n^{-1/2}S_n^* > y|\mathscr{X}],$$

where $\overline{X}$ is the mean of the sample $\mathscr{X}$. Also, for given $u < v$, define

$$S_n^{[u,v]} = \sum_{i=1}^{n}\{X_i I(u \leqslant X_i \leqslant v) - \mu\} \quad \text{and} \quad H_n(y;\, u,\, v) = P[n^{-1/2}S_n^{[u,v]} > y],$$

where $I(.)$ is the indicator function and, as before, $\mu = \mathrm{E}(X_1)$. In all the results below, $\Phi(.)$ denotes the standard normal distribution function. Our first result is the following.

Theorem 2.1. *Suppose* $\mathrm{E}(X_1) = 0$, $\mathrm{var}(X_1) = \sigma^2 > 0$, $\mu_3 = \mathrm{E}(X_1^3)$ *and* $\mathrm{E}(X_1^4) < \infty$. *Consider an arbitrary sequence* (y_n) *satisfying* $y_n \to \infty$ *such that* $y_n = o(n^{1/4})$. *Then, for* $0 \leqslant y \leqslant y_n$,

$$\hat{G}_n(y) = \left\{1 - \Phi\left(\frac{y}{\sigma}\right)\right\} \exp\left\{\frac{1}{6}n^{-1/2}\left(\frac{y}{\sigma}\right)^3\frac{\mu_3}{\sigma^3}\right\}\{1 + o_p(1)\} = H_n(y;\, X_{(1)},\, X_{(n)})\{1 + o_p(1)\}.$$

The $o_p(1)$ *terms above are both of the stated order uniformly for* $0 \leqslant y \leqslant y_n$.

Remark 2.1. This theorem tells us that when n is large, the distribution of the resample mean, conditional on $\mathscr{X}$, is close in probability to the theoretical distribution of the mean of a truncated version of the original sample, where the truncation occurs at the smallest and largest order statistics. This result is not a surprise, but we do feel that it is worth stating explicitly as it gives a clear interpretation of what, in effect, the bootstrap does.

Remark 2.2. Using arguments similar to those employed in the proof of Theorem 2.1, the following can be shown: if $\mathrm{E}(X_1^6) < \infty$ and $y_n \to \infty$, $y_n = o(n^{1/3})$, then

$$\hat{G}_n(\hat{\sigma}y) = H_n(\sigma y;\, X_{(1)},\, X_{(n)})\{1 + o_p(1)\},$$

where $\hat{\sigma}^2$ is the variance of the sample $\mathscr{X}$ and the $o_p(1)$ term is of the stated order uniformly for $0 \leqslant y \leqslant y_n$.

Theorem 2.1 says nothing about whether $\hat{G}_n(y)$ and $G_n(y)$ are close, in the sense of there being a small relative error. In the case of distributions with moment generating function finite in a neighbourhood of the origin, we have the following result, which is a slightly modified version of a theorem due to Hall (1990; 1992, Appendix V).

Theorem 2.2. *Suppose that* $\mathrm{E}(X_1) = 0$ *and that, for some* $H > 0$, $\mathrm{E}\exp(hX_1) < \infty$ *for all* $|h| < H$. *If* $y_n \to \infty$ *and* $y_n = o(n^{1/4})$, *then*

$$G_n(y) = \hat{G}_n(y)\{1 + o_p(1)\}, \tag{2.1}$$

where the $o_p(1)$ *term is of the stated order uniformly for* $0 \leqslant y \leqslant y_n$.

Proof. The proof is almost exactly the same as the proof given by Hall. The only difference is that we do not rescale by $\hat{\sigma}$ and σ. Consequently, there is a change from $y_n = o(n^{1/3})$ in Hall's result to $y_n = o(n^{1/4})$ here. Theorem 2.2 bears roughly the same relation to Hall's result that Theorem 2.1 bears to Remark 2.2. □

Our purpose now is to present results corresponding to Theorem 2.2 for three classes of sub-exponential distributions. The asymptotic behaviour of $\hat{G}_n(y)$ has already been described in broad generality by Theorem 2.1, so the principal remaining question is the asymptotic behaviour of $G_n(y)$ in sub-exponential cases.

For a review of results of this type in the Russian-language literature, see Nagaev (1979) and the references therein. We state one of these results now in a form which is convenient for later use.

Theorem 2.3 (Nagaev, 1969). *Let* $X_1, \ldots, X_n$ *be independent, identically distributed random variables with a common distribution function* F *which satisfies*

$$1 - F(x) \sim \ell(x)x^{-\nu}\{1 + o(1)\} \qquad \text{as } x \to \infty, \tag{2.2}$$

where $\ell(x)$ *is slowly-varying at infinity and* $\nu > 2$. *Suppose also that* $\mathrm{E}(X_1) = 0$, $\mathrm{var}(X_1) = \sigma^2$ *and* $\mathrm{E}|X_1|^{2+\delta} < \infty$ *for some* $\delta > 0$. *Then if* $0 < \gamma < (\nu - 2)^{1/2}$ *is fixed,*

$$G_n(y) = \{1 - \Phi(y/\sigma)\}\{1 + o(1)\},$$

where the $o(1)$ *term is of the stated order uniformly for* $1 \leqslant y \leqslant \gamma\sigma(\log n)^{1/2}$; *and if* $\gamma > (\nu - 2)^{1/2}$, *then*

$$G_n(y) = n\{1 - F(n^{1/2}y)\}\{1 + o(1)\}$$

uniformly for $y \geqslant \gamma\sigma(\log n)^{1/2}$.

This result covers our needs in the case where not all power moments are finite. The consequences for the bootstrap are as follows.

Corollary 2.1. *Suppose that the assumptions of Theorem 2.3 are satisfied with* $\nu > 4$ *(so that the assumptions of Theorem 2.1 are also satisfied), and let* y_n *be the same as in Theorem 2.1. Then if* $0 < \gamma < (\nu - 2)^{1/2}$,

$$G_n(y) = \hat{G}_n(y)\{1 + o_p(1)\} \tag{2.3}$$

uniformly for $0 \leqslant y \leqslant \sigma\gamma(\log n)^{1/2}$*; and if* $\gamma > (\nu - 2)^{1/2}$*, then*

$$\frac{G_n(y)}{\hat{G}_n(y)} \geqslant (2\pi)^{1/2} ny \exp\left\{\tfrac{1}{2}y^2(1 - \gamma^{-1})\right\}\{1 + o_p(1)\} \tag{2.4}$$

uniformly for $\sigma\gamma\,(\log n)^{1/2} \leqslant y \leqslant y_n$.

Proof. Most of this result follows directly from Theorems 2.1 and 2.3. The particular form which appears on the right of (2.4) follows from an argument rather similar to that used to prove part (ii) of Theorem 2.4 below; we omit the details. □

The following three remarks are directed at Corollary 2.1; they apply equally to Corollaries 2.2 and 2.3 below.

Remark 2.3. It is worth emphasizing that, under the conditions of Theorem 2.3 and Corollary 2.1, the approximation of $G_n(y)$ by $\hat{G}_n(y)$ breaks down precisely when the normal approximation to $G_n(y)$ breaks down. Note that the breakdown of the normal approximation is quite dramatic.

Remark 2.4. Corollary 2.1 does not tell us what happens when $y > y_n$. However, it seems likely that the bootstrap approximation to $G_n(y)$ will continue to deteriorate as y increases beyond y_n, though more elaborate arguments than those given in this paper would be needed to prove this rigorously.

Remark 2.5. Investigation of the relative error properties of more general statistics (e.g. a studentized mean) would certainly be of interest, as would the study of the relative coverage error of the various types of bootstrap confidence interval.

Note that the lognormal distribution has finite power moments and therefore does not satisfy (2.2). Our next task is to present results roughly along the lines of Theorem 2.3, but which cover sub-exponential distributions, such as the lognormal, whose power moments are all finite.

First, we introduce some assumptions.

(F1) As $x \to \infty$,

$$1 - F(x) \sim \exp\{-g(x)\}\{1 + o(1)\}$$

where $g(x)$ satisfies the following conditions (as $x \to \infty$): (a) $g(x)$ is ultimately strictly increasing and $g(x) \to \infty$; (b) $x^{-1}g(x)$ is ultimately strictly decreasing and $x^{-1}g(x) \to 0$; and (c) $\{\log x\}^{-1}g(x) \to \infty$.

(F2) The function g in (F1) is slowly varying at infinity; and $\mathrm{E}|X_1|^4 < \infty$.

(F3) The function g in (F1) is regularly varying at infinity with coefficient of variation $\alpha \in (0, 1)$, i.e. for any $\rho > 0$, $g(\rho x)/g(x) \to \rho^\alpha$ as $x \to \infty$; and $\mathrm{E}|X_1|^r < \infty$ for some $r > \max\{4, 2(1 - \alpha)^{-1}\}$.

Some brief comments on assumptions (F1)–(F3) now follow. In (F1), (a) and (b) are mild technical conditions. Condition (b) also ensures that $\int e^{hx}\,dF(x)=\infty$ for all $h>0$, which rules out the case covered by Theorem 2.2. The purpose of (c) is to rule out the case covered by Theorem 2.3. Conditions (F2) and (F3) are mutually exclusive; it is helpful to formulate theorems for the cases covered by (F2) and (F3) separately even though the proofs are rather similar.

In Theorems 2.4 and 2.5 below, we shall require the following definitions. For $b<a$ define the integer moments

$$\mu_r(b, a)=\int_{[b,a]}|x|^r\,dF(x)$$

of $X_1 I(b\leqslant X_1\leqslant a)$, and write

$$\Delta_F(b, a)=\sup_{r\geqslant 1}\left\{\frac{\mu_{r+4}(b, a)}{r}\right\}^{1/r},$$

where the supremum is taken over integer values of r. As far as we are aware, for all sub-exponential distributions of practical interest which satisfy (F1), the following holds:

$$\Delta_F(0, a)\sim\frac{a}{g(a)}\qquad\text{as } a\to\infty, \tag{2.5}$$

where F and g are related via (F1). In the theorems below, we prefer to assume (2.5) and then check it on a case-by-case basis as in the examples of Section 3.

Let λ_n be the solution in $y\geqslant 1$ of

$$\frac{y^2}{2\sigma^2}=g(n^{1/2}y), \tag{2.6}$$

where $\sigma^2=\mathrm{var}(X_1)$. Note that, as a consequence of (F1), $\lambda_n\geqslant 1$ exists and is unique when n is sufficiently large.

Theorem 2.4. *Let F be the distribution function of a random variable with mean zero and variance σ^2. If, in addition, (F1), (F2) and (2.5) are satisfied, then the following statements hold.*

(i) For any fixed $\gamma\in(0, 1)$ and $0\leqslant y\leqslant\gamma\lambda_n$,

$$G_n(y)=\left\{1-\Phi\left(\frac{y}{\sigma}\right)\right\}\{1+o(1)\}, \tag{2.7}$$

where the $o(1)$ remainder term is of the stated order uniformly for $0\leqslant y\leqslant\gamma\lambda_n$.

(ii) For any fixed $\gamma>1$ and $y\geqslant\gamma\lambda_n$,

$$\frac{G_n(y)}{1-\Phi(y/\sigma)}\geqslant(\pi\sigma^2/2)^{1/2}ny\exp\left\{\frac{(1-\gamma^{-1})y^2}{2\sigma^2}\right\}\{1+o(1)\} \tag{2.8}$$

uniformly for $y\geqslant\gamma\lambda_n$.

The proof is postponed until Section 5.

Theorems 2.1 and 2.4 combine to produce the following analogue of Theorem 2.2.

Corollary 2.2. *Suppose that the assumptions of Theorem 2.4 (and therefore the assumptions of Theorem 2.1) are satisfied, and that y_n is as before. Then we have the following.*

(i) For any fixed $\gamma \in (0, 1)$,

$$G_n(y) = \hat{G}_n(y)\{1 + o_p(1)\}$$

uniformly for $0 \leqslant y \leqslant \gamma\lambda_n$.

(ii) If $\gamma > 1$ then

$$\frac{G_n(y)}{\hat{G}_n(y)} \geqslant (\pi\sigma^2/2)^{1/2} n y \exp\left\{\frac{y^2(1-\gamma^{-1})}{2\sigma^2}\right\}\{1 + o_p(1)\}$$

uniformly for $\gamma\lambda_n \leqslant y \leqslant y_n$.

Proof. Follows directly from Theorem 2.1 and Theorem 2.4. □

We now describe what happens when condition (F2) is replaced by condition (F3). Nagaev (1973; 1979, Theorem 2.1) has proved a theorem which is directly relevant to distributions satisfying (F3). However, both the statement and the proof of this result are very complex. We prefer to give a much simpler, but less complete, result which is tailored to the present objective of describing the breakdown of the normal approximation.

Theorem 2.5. *Let F be a distribution function which satisfies all the assumptions of Theorem 2.4, but with (F2) replaced by (F3). Suppose that (y_n) is any sequence satisfying $y_n \gg n^{(1/4)-\varepsilon}$ for any $\varepsilon > 0$, and $y_n = o(n^{1/4})$.*

(i) If $0 < \alpha < 2/3$, then for any fixed $0 < \gamma < 2^{-1/(2-\alpha)}$ and $0 \leqslant y \leqslant \gamma\lambda_n$,

$$\frac{G_n(y)}{1 - \Phi(y/\sigma)} = \exp\left\{\frac{1}{6} n^{-1/2}\left(\frac{y}{\sigma}\right)^3 \frac{\mu_3}{\sigma_3}\right\}\{1 + o(1)\} \tag{2.9}$$

uniformly for $0 \leqslant y \leqslant \gamma\lambda_n$.

(ii) If $0 < \alpha < \frac{2}{3}$ then for any fixed $\gamma > 1$ and $y \in [\gamma\lambda_n, y_n]$,

$$\frac{G_n(y)}{\{1 - \Phi(y/\sigma)\}\exp\left\{\frac{1}{6} n^{-1/2}(y/\sigma)^3 \frac{\mu_3}{\sigma^3}\right\}} \geqslant (\pi\sigma^2/2)^{1/2} n y \exp\left\{\frac{(1-\gamma^{-1})y^2}{2\sigma^2}\right\}\{1 + o(1)\} \tag{2.10}$$

uniformly for $y \in [\gamma\lambda_n, y_n]$.

(iii) If $\alpha > \frac{2}{3}$ then for $0 \leqslant y \leqslant y_n$,

$$G_n(y) = \left\{1 - \Phi\left(\frac{y}{\sigma}\right)\right\} \exp\left\{\frac{1}{6} n^{-1/2} \left(\frac{y}{\sigma}\right)^3 \frac{\mu_3}{\sigma^3}\right\}\{1 + o(1)\} \tag{2.11}$$

uniformly for $0 \leqslant y \leqslant y_n$.

The proof is given in Section 5.

The analogue of Corollary 2.2 in this case is as follows.

Corollary 2.3. *Suppose the assumptions of Theorem 2.5 (and therefore the assumptions of Theorem 2.1) are satisfied. Then parts (i)–(iii) of Theorem 2.5 hold with the following changes: in (2.9)–(2.11),* $\hat{G}_n(y)$ *replaces*

$$\left\{1 - \Phi\left(\frac{y}{\sigma}\right)\right\} \exp\left\{\frac{1}{6} n^{-1/2} \left(\frac{y}{\sigma}\right)^3 \frac{\mu_3}{\sigma^3}\right\};$$

and the $o(1)$ *terms in (2.9)–(2.11) are replaced by* $o_p(1)$ *terms, where the latter inherit the uniformity properties (in probability) of the former in each case.*

Proof. Follows directly from Theorem 2.1 and Theorem 2.5. □

3. Examples

We now apply the results of the previous section to several examples including the lognormal.

3.1. The lognormal case

Suppose that $X = \exp\{\mathrm{N}(0, \tau^2)\}$. Then the tail probability $P[X > x]$ as $x \to \infty$ is given by

$$P[X > x] \sim \exp\{-g(x)\},$$

where we may write $g(x) = g_0(x) + g_1(x)$ with

$$g_0(x) = (2\tau^2)^{-1}(\log x)^2 \quad \text{and} \quad g_1(x) = \log\log x + \tfrac{1}{2}\log\left(\frac{2\pi}{\tau^2}\right).$$

It is easy to check that g satisfies conditions (F1) and (F2).

Let $\sigma^2 = \mathrm{var}(X) = \mathrm{e}^{2\tau^2} - \mathrm{e}^{\tau^2}$ and, as before, write λ_n for the solution in $x \geqslant 1$ of

$$\frac{1}{2\sigma^2} x^2 = g(n^{1/2} x).$$

Condition (F1) implies that, when n is sufficiently large, λ_n exists and is unique. Since $g_1(x) = o\{g_0(x)\}$ as $x \to \infty$, it follows that

$$\lambda_n \sim \frac{\sigma}{2\tau} \log n \tag{3.1}$$

as $n \to \infty$.

In order to apply Theorem 2.4, we need to check that $\Delta_F(0, a) \sim a/g(a)$ as $a \to \infty$, where $\Delta_F(0, a)$ is defined before Theorem 2.4. Using elementary calculations,

$$\mu_r(0, a) = \mathrm{E}[X^r I(X \leqslant a)] = \mathrm{e}^{r^2\tau^2/2}\Phi\left(\frac{\log a - r\tau^2}{\tau}\right), \tag{3.2}$$

where Φ is the standard normal distribution function and, below, ϕ is the standard normal density. Using well-known properties of Mills's ratio $\{1-\Phi\}/\phi$, we obtain the following crude bounds on $\mu_r(0, a)$:

$$\mu_r(0, a) \leqslant \begin{cases} \mathrm{e}^{r^2\tau^2/2} & \text{if } r \leqslant \tau^{-2}(\tau + \log a), \\ C_1 \mathrm{e}^{r\log a - g(a)} & \text{if } r > \tau^{-2}(\tau + \log a), \end{cases}$$

where, as $a \to \infty$, C_1 does not depend on r or a. Using Stirling's approximation, we have

$$\{\Gamma(r+1)\}^{1/r} = \mathrm{e}^{-1} r\{1 + \varepsilon(r)\} \tag{3.3}$$

where $\varepsilon(r) \to 0$ as $r \to \infty$. Consequently, when a is sufficiently large and $r \geqslant 1$ is an integer,

$$\left\{\frac{\mu_{r+4}(0, a)}{r!}\right\}^{1/r} \leqslant C_2 a^{1/2} \tag{3.4}$$

if $r + 4 \leqslant \tau^{-2}(\tau + \log a)$; and if $r + 4 > \tau^{-2}(\tau + \log a)$, then

$$\left\{\frac{\mu_{r+4}(0, a)}{r!}\right\}^{1/r} \leqslant \eta(r) r^{-1} \exp\left\{\log a - \frac{g(a) - 4\log a}{r}\right\}. \tag{3.5}$$

In (3.4), C_2 stays bounded as $a \to \infty$ and in (3.5), $\eta(r) = C_1^{1/r}\{1 + \varepsilon(r)\}^{-1} \to \mathrm{e}$ as $r \to \infty$. It is straightforward to show that as, $a \to \infty$, (3.5) has an upper bound of size

$$\frac{a}{g(a)}\{1 + o(1)\},$$

since $g(a) \sim g(a) - 4\log a$. Thus

$$\Delta_F(0, a) \leqslant \max\left\{\mathrm{C}_2 a^{1/2}, \frac{a}{g(a)}\{1 + o(1)\}\right\} \sim \frac{a}{g(a)}\{1 + o(1)\}.$$

To show that this bound is achieved asymptotically, we evaluate $\{\mu_{r+4}(0, a)\}^{1/r}$ with r equal to the integer part of $g(a)$, using (3.2) and (3.3). The details are straightforward.

Thus we may apply Theorems 2.1 and 2.4 and Corollary 2.2, and it is found that the normal approximation $1 - \Phi(y/\sigma)$ to $G_n(y)$, and the bootstrap approximation $\hat{G}_n(y)$ to $G_n(y)$ both break down around $y = \lambda_n$, with the asymptotic behaviour of λ_n described in (3.1).

3.2. The Weibull case

Consider the Weibull distribution, $Wei(\alpha, c)$, with density

$$f(x) = c\alpha x^{\alpha-1}\exp(-cx^\alpha), \qquad \alpha, c, x > 0, \tag{3.6}$$

and distribution function $F(x) = 1 - \exp(-cx^\alpha)$. If $\alpha \geqslant 1$ then the moment generating function is finite in a neighbourhood of the origin and we are in the situation covered by Theorem 2.2. We shall focus on the case $\alpha \in (0, 1)$.

When $\alpha \in (0, 1)$ it is easy to check that conditions (F1) and (F3) are satisfied with $g(x) = cx^\alpha$. Using both one-sided and two-sided versions of Laplace's approximation, it can be shown that

$$\Delta_F(0, a) \sim \frac{a}{g(a)} \qquad \text{as } a \to \infty. \tag{3.7}$$

We shall not give a derivation of (3.7) here but we just mention that the simplest approach is to follow the two-step strategy followed in the lognormal case: (i) show, using relatively crude approximations that $\Delta_F(0, a) \leqslant \{a/g(a)\}\{1 + o(1)\}$; and (ii) show that this bound is achieved asymptotically by the 'optimal' choice of r.

In the Weibull case, it is easily seen that λ_n is given by

$$\lambda_n = (2c\sigma^2)^{1/(2-\alpha)} n^{\alpha/(4-2\alpha)}$$

where

$$\sigma^2 = \mathrm{var}\{Wei(\alpha, c)\} = c^{-2/\alpha}\left\{\Gamma\left(\frac{\alpha+2}{\alpha}\right) - \Gamma\left(\frac{\alpha+1}{\alpha}\right)^2\right\}.$$

If $0 < \alpha < \frac{2}{3}$, we are in the situation covered by parts (i) and (ii) of Theorem 2.5 and Corollary 2.3; if $\frac{2}{3} < \alpha < 1$ we are in the situation of part (iii) of Theorem 2.5 and Corollary 2.3; and if $\alpha \geqslant 1$, we are in the situation covered by Theorem 2.2.

3.3. The log-Weibull case

If $X = \exp\{Wei(\alpha, c)\}$, then X is said to have a log-Weibull distribution (cf. the definition of the lognormal). We shall restrict attention to the case in which the index of the Weibull distribution (α in (3.6)) satisfies $\alpha > 1$. Then all moments are finite but the moment generating function is not finite in any neighbourhood of the origin. The appropriate choice for g is $g(x) = c|\log x|^\alpha$. It is straightforward to check that conditions (F1) and (F2) are satisfied. Moreover, as in the other examples, it can be shown that $\Delta_F(0, a) \sim a/g(a)$ as $a \to \infty$; we omit the details. Thus we are back in the domain of Theorem 2.4. In this case, λ_n is given by

$$\lambda_n \sim 2^{(1-\alpha)/2} c^{1/2} \sigma (\log n)^{\alpha/2}$$

as $n \to \infty$, where now $\sigma^2 = \mathrm{var}(\mathrm{e}^{Wei(\alpha, c)})$. Note that, as one might expect, when $\alpha = 2$, λ_n grows at a similar rate to the lognormal case.

3.4. The *t* distribution

Let T be a random variable which has a t distribution with ν degrees of freedom. The density of T is given by

$$f_\nu(x) = c_\nu \nu^{-1/2}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2},$$

where $c_\nu = \Gamma\{(\nu+1)/2\}/\{\Gamma(\nu/2)\Gamma(1/2)\}$. As $x \to \infty$, the tail probability $P[T > x]$ satisfies

$$P[T > x] \sim \exp\{-g(x)\},$$

where

$$g(x) = \nu \log x - \log c_\nu - \frac{\nu-2}{2}\log \nu.$$

The fourth moment of T exists if $\nu > 4$, in which case we may apply Theorem 2.3 and Corollary 2.1 to show that $\lambda_n \sim (\nu-2)^{1/2}\sigma(\log n)^{1/2}$.

It is interesting to note that, in the case of the t distribution, $\Delta_F(0, a) \sim \theta a/g(a)$ where $\theta = \nu/(\nu-4) > 1$; we omit the derivation. It seems that, in general, (2.5) does not hold for sub-exponential distributions unless all power moments are finite.

4. Auxiliary results

We begin with an elementary result whose proof is included for completeness.

Lemma 4.1. *Let $X_1, \ldots, X_n$ be independent, identically distributed random variables and suppose that $\mathrm{E}|X_1|^r < \infty$ for some $r > 0$. Then for any $0 < s \leqslant r$,*

$$P[X_{(n)} > n^{1/s}] = o(n^{-(r-s)/s}) \qquad \text{as } n \to \infty.$$

Moreover, $n^{-1/r}X_{(n)} \to 0$ in probability.

Proof. From the definition of order statistics,

$$P[X_{(n)} \leqslant n^{1/s}] = \{1 - P[X_1 > n^{1/s}]\}^n,$$

and, by a Markov–Chebyshev argument,

$$\{P[X_1 \leqslant n^{1/s}]\}^n \geqslant \{1 - n^{-1}n^{-(r-s)/s}\mathrm{E}[|X_1|^r I(X_1 > n^{1/s})]\}^n$$

$$= 1 - o(n^{-(r-s)/s}),$$

since $\mathrm{E}[|X_1|^r I(X_1 > n^{1/s})] \to 0$ as $n \to \infty$ by the dominated convergence theorem. Therefore $P[X_{(n)} > n^{1/s}] = o(n^{-(r-s)/s})$. The second part follows easily from the proof of the first part. □

Our next result plays a key rôle in the proofs of Theorem 2.4 and 2.5. See Hahn and Klass (1997) and the references therein for discussion of related results.

Lemma 4.2. *Suppose that* $\mathrm{E}(X_1) = 0$ *and* $\mathrm{E}|X_1|^r < \infty$ *for some* $r > 2$. *Choose* $y_0 > 0$ *and* $\theta \in (0, 1]$, *and write* $b = b_n = -n^{1/r}$ *and* $a = a_n = \theta n^{1/2} y$. *Then for* $y \geqslant y_0$,

$$P[n^{-1/2} S_n > y] = P[n^{-1/2} S_n^{[b,a]} > y]\{1 + R_{1,n}(y)\} + J_n(y),$$

where

$$\{\tfrac{1}{2} + R_{2,n}(y)\} P[X_{(n)} > n^{1/2} y] \leqslant J_n(y) \leqslant \{1 + R_{3,n}(y)\} P[X_{(n)} > a]$$

and, as $n \to \infty$,

$$\sup_{y \geqslant y_0} |R_{i,n}(y)| = o(1), \qquad i = 1, 2, 3.$$

Proof. Consider the following identities:

$$\begin{aligned} P[n^{-1/2} S_n > y] = {} & P[n^{-1/2} S_n^{[b,a]} > y, X_{(1)} \geqslant b, X_{(n)} \leqslant a] \\ & + P[n^{-1/2} S_n > y, X_{(1)} < b, X_{(n)} \leqslant a] \\ & + P[n^{-1/2} S_n > y, X_{(n)} > a]; \end{aligned} \tag{4.1}$$

$$\begin{aligned} & P[n^{-1/2} S_n^{[b,a]} > y, X_{(1)} \geqslant b, X_{(n)} \leqslant a] \\ & \quad = P[n^{-1/2} S_n^{[b,a]} > y] + P[n^{-1/2} S_n^{[b,a]} > y, X_{(1)} < b, X_{(n)} > a] \\ & \quad\quad - P[n^{-1/2} S_n^{[b,a]} > y, X_{(1)} < b] - P[n^{-1/2} S_n^{[b,a]} > y, X_{(n)} > a]; \end{aligned} \tag{4.2}$$

$$\begin{aligned} & P[n^{-1/2} S_n > y, X_{(1)} < b, X_{(n)} \leqslant a] \\ & \quad = P[n^{-1/2} S_n > y, X_{(1)} < b] - P[n^{-1/2} S_n > y, X_{(1)} < b, X_{(n)} > a]; \end{aligned} \tag{4.3}$$

$$P[n^{-1/2} S_n > y, X_{(n)} > a] = P[X_{(n)} > a] - P[n^{-1/2} S_n \leqslant y, X_{(n)} > a]. \tag{4.4}$$

First, note that the second term on the right of (4.2) and the second term on the right of (4.3) are both bounded above by $P[X_{(1)} < b, X_{(n)} > a]$. If F is the distribution function of X_1 then

$$P[X_{(1)} < b] = 1 - \{1 - F(b-)\}^n, \qquad P[X_{(n)} > a] = 1 - F(a)^n$$

and

$$P[X_{(1)} < b, X_{(n)} > a] = 1 - \{1 - F(b-)\}^n - F(a)^n - \{F(a) - F(b-)\}^n.$$

But Lemma 4.1 tells us that, with the given choices of a and b, $P[X_{(1)} < b]$ and $P[X_{(n)} > a]$ both converge to zero. Consequently, $n\{1 - F(a)\}$ and $nF(b-)$ both converge to zero and

$$P[X_{(1)} < b,\, X_{(n)} > a] \sim n^2 F(b-)\{1 - F(a)\}$$
$$\sim P[X_{(1)} < b]P[X_{(n)} > a]$$
$$= o\{P[X_{(n)} > a]\} \qquad \text{as } n \to \infty. \tag{4.5}$$

Let us now consider the second term on the right of (4.4). We have

$$P[n^{-1/2}S_n \leqslant y,\, X_{(n)} > a] = P[n^{-1/2}S_n \leqslant y,\, X_{(n)} \in (a,\, n^{1/2}y]]$$
$$+ P[n^{-1/2}S_n \leqslant y,\, X_{(n)} > n^{1/2}y]. \tag{4.6}$$

But

$$P[n^{-1/2}S_n \leqslant y,\, X_{(n)} \in (a,\, n^{1/2}y]] \leqslant P[X_{(n)} > a] - P[X_{(n)} > n^{1/2}y] \tag{4.7}$$

and

$$P[n^{-1/2}S_n \leqslant y,\, X_{(n)} > n^{1/2}y] = \int_{(n^{1/2}y,\infty)} P[n^{-1/2}S_n \leqslant y | X_{(n)} = z]\, \mathrm{d}F(z)^n, \tag{4.8}$$

where $P[n^{-1/2}S_n \leqslant y | X_{(n)} = z]$ may be arbitrarily defined as zero if z lies outside the support of the distribution of X_1. Elementary properties of order statistics imply that

$$P[n^{-1/2}S_n \leqslant y | X_{(n)} = z] \leqslant P[n^{-1/2}S_{n-1}^{(-\infty,z)} + n^{-1/2}z \leqslant y]$$

with equality if z lies in the support of X_1 and also X_1 does not have an atom at z. For $z > n^{1/2}y$,

$$P[n^{-1/2}S_{n-1}^{(-\infty,z)} + n^{-1/2}z \leqslant y] \leqslant P[n^{-1/2}S_{n-1}^{(-\infty,n^{1/2}y]} \leqslant 0]$$
$$\leqslant P[(n-1)^{-1/2}S_{n-1} \leqslant 0,\, X_{(n-1)} \leqslant n^{1/2}y]$$
$$+ P[X_{(n-1)} > n^{1/2}y]$$
$$\leqslant P[(n-1)^{-1/2}S_{n-1} \leqslant 0] + P[X_{(n-1)} > n^{1/2}y]$$
$$= \tfrac{1}{2} + o(1) \tag{4.9}$$

where in (4.9), but not elsewhere, $X_{(n-1)}$ is the largest order statistic in a sample of size $n-1$ from F. Consequently, using (4.7)–(4.9),

$$P[n^{-1/2}S_n \leqslant y,\, X_{(n)} > a] \leqslant P[X_{(n)} > a] - P[X_{(n)} > n^{1/2}y]$$
$$+ \{\tfrac{1}{2} + o(1)\}P[X_{(n)} > n^{1/2}y],$$

and, returning to (4.4),

$$P[n^{-1/2}S_n > y,\, X_{(n)} > a] \geqslant \{\tfrac{1}{2} + o(1)\}P[X_{(n)} > n^{1/2}y]. \tag{4.10}$$

Consider now the third and fourth terms on the right of (4.2). We have

$$P[n^{-1/2}S_n^{[b,a]} > y,\, X_{(1)} < b] = P[n^{-1/2}S_n^{[b,a]} > y | X_{(1)} < b] P[X_{(1)} < b]$$

where, by elementary calculation,

$$P[n^{-1/2}S_n^{[b,a]} > y | X_{(1)} < b] = \frac{\sum_{r=1}^{n-1} C_{r,n}\pi^r(1-\pi)^{n-r} P[n^{-1/2}S_{n-r}^{[b,a]} > y]}{1-(1-\pi)^n} \tag{4.11}$$

where $C_{r,n} = n!/\{r!(n-r)!\}$ and $\pi = P[X_1 < b] = F(b-)$. Since (i) $n\pi \to 0$ as $n \to \infty$ and (ii) $P[X_1 \geqslant 0] > 0$, because $\mathrm{E}(X_1) = 0$ and $\mathrm{var}(X_1) > 0$, it follows from (4.11) that

$$P[n^{-1/2}S_n^{[b,a]} > y | X_{(1)} < b] \sim P[n^{-1/2}S_{n-1}^{[b,a]} > y].$$

But

$$P[n^{-1/2}S_n^{[b,a]} > y] \geqslant P[n^{-1/2}S_{n-1}^{[b,a]} > y] P[X_n \geqslant 0]$$

and therefore

$$\begin{aligned} P[n^{-1/2}S_n^{[b,a]} > y,\, X_{(1)} < b] &= O\{P[n^{-1/2}S_n^{[b,a]} > y] P[X_{(1)} < b]\} \\ &= o\{P[n^{-1/2}S_n^{[b,a]} > y]\}. \end{aligned} \tag{4.12}$$

An identical argument shows that

$$P[n^{-1/2}S_n^{[b,a]} > y,\, X_{(n)} > a] = o\{P[n^{-1/2}S_n^{[b,a]} > y]\}. \tag{4.13}$$

Next, consider the first term on the right of (4.3). We have

$$P[n^{-1/2}S_n > y,\, X_{(1)} < b] = \int_{(-\infty,b)} P[n^{-1/2}S_n > y | X_{(1)} = z] \, \mathrm{d}F(z)^n \tag{4.14}$$

where, for $z < b$,

$$\begin{aligned} P[n^{-1/2}S_n > y | X_{(1)} = z] &\leqslant P[n^{-1/2}S_{n-1}^{(z,\infty)} + n^{-1/2}z > y] \\ &\leqslant P[n^{-1/2}S_{n-1}^{[b,\infty)} + n^{-1/2}b > y] \\ &\leqslant P[n^{-1/2}S_n^{[b,\infty)} > y], \end{aligned}$$

and therefore, using (4.14),

$$P[n^{-1/2}S_n > y,\, X_{(1)} < b] \leqslant P[n^{-1/2}S_n^{[b,\infty)} > y] P[X_{(1)} < b]. \tag{4.15}$$

But

$$P[n^{-1/2}S_n^{[b,\infty)} > y] \leqslant P[n^{-1/2}S_n^{[b,a]} > y] + P[X_{(n)}^{[b,\infty)} > a], \tag{4.16}$$

where $X_{(n)}^{[b,\infty)}$ is the largest of $X_1 I(X_1 \geqslant b), \ldots, X_n I(X_n \geqslant b)$, and

$$
\begin{aligned}
P[X_{(n)}^{[b,\infty)} > a] &= 1 - \left(\frac{F(a) - \pi}{1 - \pi}\right)^n \\
&\sim n\{1 - F(a)\} \\
&\sim P[X_{(n)} > a], \qquad (4.17)
\end{aligned}
$$

with $\pi = P[X_1 < b]$ as before. Therefore, using (4.15)–(4.17),

$$P[n^{-1/2}S_n > y,\ X_{(1)} < b] = o\{P[n^{-1/2}S_n^{[b,a]} > y]\} + o\{P[X_{(n)} > a]\}. \qquad (4.18)$$

Finally, the desired conclusion follows after substituting (4.5), (4.10), (4.12), (4.13) and (4.18) into (4.2)–(4.4), and then using (4.1). □

Let us now recall a result due to Petrov (1995, Theorem 5.23); see also Petrov (1975). We shall give a slightly different, but equivalent, statement of the result.

Theorem 4.1. *Let $X_1, \ldots, X_n$ be an sample of independent, zero-mean random variables from a non-degenerate distribution whose moment generating function is finite in a neighbourhood of the origin, i.e. for some $H > 0$, $R(h) = \mathrm{E}\exp(hX_1) < \infty$ for all $|h| < H$. Write $S_n = \sum_{i=1}^n X_i$ and $\sigma^2 = \mathrm{var}(X_1)$. Then if $y_n \to \infty$, $y_n = o(n^{1/2})$, we have*

$$\frac{P[n^{-1/2}S_n > y]}{1 - \Phi(y/\sigma)} = \exp\left\{n^{-1/2}\left(\frac{y}{\sigma}\right)^3 \lambda\left(\frac{y}{\sigma n^{1/2}}\right)\right\}\left\{1 + O\left(\frac{y+1}{n^{1/2}}\right)\right\}, \qquad (4.19)$$

with $\lambda(t)$ defined as follows:

$$t^3\lambda(t) = \tfrac{1}{2}t^2 + \log R(\hat{h}) - \hat{h}\sigma t \qquad (4.20)$$

where $\hat{h} = \hat{h}(t)$ is the unique solution of $d\log R(h)/dh = \sigma t$.

Proof. See Petrov (1995). □

We shall consider two variations of Petrov's result which can be applied to triangular arrays but, for simplicity, we shall restrict attention to deviations of size $o(n^{1/4})$ rather that $o(n^{1/2})$. Let $(F^{(n)})_{n\geqslant 1}$ be a sequence of distribution functions. For each n, let $X_1^{(n)}, \ldots, X_n^{(n)}$ be a random sample from $F^{(n)}$. Write $\mu^{(n)} = \mathrm{E}(X_1^{(n)})$, $\sigma^{(n)2} = \mathrm{var}(X_1^{(n)})$, $\mu_r^{(n)} = \mathrm{E}|X_1^{(n)} - \mu^{(n)}|^r$, $r = 3, 4, \ldots,$ and consider the following conditions.

(C1) For some sequence $(d_n)_{n\geqslant 1}$ satisfying $d_n \to \infty$, $|X_i^{(n)}| \leqslant d_n$ with probability one.
(C2) For some constant $1 < A < \infty$,

$$\liminf_{n\geqslant 1} \sigma^{(n)} > A^{-1} \quad \text{and} \quad \limsup \mu_4^{(n)} < A.$$

Write $S_n^{(n)} = \sum_{i=1}^n (X_i^{(n)} - \mu^{(n)})$.

Proposition 4.1. *Let $(F^{(n)})_{n\geqslant 1}$ be a sequence of zero-mean distribution functions which satisfy (C1) and (C2). Fix any $\gamma \geqslant A$. Suppose that the sequence (y_n) satisfies $y_n \to \infty$,*

$y_n = o(n^{1/4})$ and $d_n y_n \leqslant n^{1/2}\gamma$ when n. Then

$$\frac{P[n^{-1/2}S_n^{(n)} > y]}{1 - \Phi(y/\sigma^{(n)})} = \exp\left\{\frac{1}{6}n^{-1/2}\left(\frac{y}{\sigma^{(n)}}\right)^3 \frac{\mu_3^{(n)}}{\sigma^{(n)3}}\right\}\{1 + R_n(y)\}$$

where

$$|R_n(y)| \leqslant C_0\{n^{-1}y^4 + n^{-1/2}(y+1) \qquad \text{for } 0 \leqslant y \leqslant y_n, \tag{4.21}$$

and $C_0 = C_0(\gamma)$, which depends only on γ, is finite for all $\gamma < \infty$.

Remark 4.1. In our principal applications of Proposition 4.1, $(F^{(n)})$ will represent a (random) sequence of distribution functions depending on samples of size n from a fixed underlying distribution. In such cases, d_n, $\sigma^{(n)}$ and $\mu_4^{(n)}$ will be random. With very minor modifications to the proof of the deterministic version of Proposition 4.1 given below, it can be shown that

$$\frac{P[n^{-1/2}S_n^{(n)} > y|\mathscr{X}]}{1 - \Phi(y/\sigma^{(n)})} = \exp\left\{\frac{1}{6}n^{-1/2}\left(\frac{y}{\sigma^{(n)}}\right)^3 \frac{\mu_3^{(n)}}{\sigma^{(n)3}}\right\}\{1 + o_p(1)\}$$

uniformly for $0 \leqslant y \leqslant y_n$, given the following: (i) $y_n \to \infty$ and $y_n = o(n^{1/4})$; (ii) $d_n = o_p(n^{1/4})$; and (iii) for some constant $0 < A < \infty$, $P[\sigma^{(n)} < A^{-1}] \to 0$ and $P[\mu_4^{(n)} > A] \to 0$.

We shall also require a different version of Proposition 4.1. Consider the following alternative to condition (C1):

(C1′) For all integers $r \geqslant 4$ and $n \geqslant 1$, $\mu_{r+4}^{(n)} \leqslant r!d_n^r$.

Note that in (C1′) there is no assumption that $X_i^{(n)}$ has compact support.

Proposition 4.2. *Let $(F^{(n)})_{n\geqslant 1}$ be a sequence of zero-mean distribution functions which satisfy* (C1′) *and* (C2). *Fix any γ such that $(1 - A^{-1}) \leqslant \gamma < 1$, and suppose that $y_n \to \infty$, $y_n = o(n^{1/4})$ and $d_n y_n \leqslant n^{1/2}\sigma^{(n)2}\gamma$ when n is sufficiently large. Then the conclusion of Proposition 4.1 holds, but possibly with C_0 in (4.21) replaced by $C_1 = C_1(\gamma)$ which depends only on γ and is finite for all $\gamma < 1$.*

Proof of Proposition 4.1. First, note that $|X_i^{(n)}| \leqslant d_n$ implies that $|\mu_{r+4}^{(n)}| \leqslant 2^r d_n^r \mu_4^{(n)}$, for $r = 0, 1, \ldots$. Write

$$R(h) = R^{(n)}(h) = \mathrm{E}\exp\{h(X_1^{(n)} - \mu^{(n)})\}$$

and, to simplify notation, drop the superscript (n) in $\sigma^{(n)}$ and $\mu_r^{(n)}$, $r \geqslant 3$. Then for any h satisfying

$$h \geqslant 0, \qquad d_n h \leqslant 2\gamma \quad \text{and} \quad \mu_4 \leqslant \gamma, \tag{4.22}$$

$$\begin{aligned} R(h) &= 1 + \frac{1}{2!}\sigma^2 h^2 + \frac{1}{3!}\mu_3 h^3 + \sum_{r=0}^{\infty} \frac{1}{(r+4)!}\mu_{r+4} h^{r+4} \\ &= 1 + \frac{1}{2!}\sigma^2 h^2 + \frac{1}{3!}\mu_3 h^3 + O\left(\frac{1}{4!}\mu_4 h^4 \sum_{r=0}^{\infty} \frac{4!}{(r+4)!} 2^r d_n^r h^r\right) \\ &= 1 + \frac{1}{2!}\sigma^2 h^2 + \frac{1}{3!}\mu_3 h^3 + O\left(\frac{1}{4!}\mu_4 h^4 \exp(4\gamma)\right) \\ &= 1 + \frac{1}{2!}\sigma^2 h^2 + \frac{1}{3!}\mu_3 h^3 + O(h^4) \end{aligned} \tag{4.23}$$

in view of (4.21). Similarly, still assuming that h satisfies (4.22), we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}h} R(h) &= \sigma^2 h + \frac{1}{2!}\mu_3 h^2 + O(h^3), \\ \frac{\mathrm{d}^2}{\mathrm{d}h^2} R(h) &= \sigma^2 + \mu_3 h + O(h^2), \\ \frac{\mathrm{d}^3}{\mathrm{d}h^3} R(h) &= \mu_3 + O(h), \\ \frac{\mathrm{d}^4}{\mathrm{d}h^4} R(h) &= O(1), \end{aligned}$$

where in each case the absolute value of the $O(h^k)$ remainder term is bounded above by an expression of the form $L(\gamma)h^k$ where, here and below, $L(\gamma)$ is used generically for a function which depends only on γ and is finite for all $\gamma > 0$.

Define $m(h) = \{R(h)\}^{-1}\mathrm{d}R(h)/\mathrm{d}h$. For h satisfying (4.22),

$$m(h) = \sigma^2 h + \frac{1}{2!}\mu_3 h^2 + O(h^3),$$

where the $O(h^3)$ remainder term is bounded by $L(\gamma)h^3$. If n is sufficiently large (so that, by (C1), d_n is large) and $t \geqslant 0$ satisfies $t \leqslant \sigma\gamma/d_n$, the unique solution $h = \hat{h}$ of $m(h) = \sigma t$ is of the form

$$\hat{h} = \frac{t}{\sigma} - \frac{1}{2\sigma}\mu_3 \sigma^3 t^2 + O(t^3) \qquad \text{as } t \to 0, \tag{4.24}$$

and the absolute value of the remainder term $O(t^3)$ in (4.23) is bounded by $L(\gamma)t^3$. For such t we find, after substituting (4.24) into (4.20) using (4.15), that

$$t^3\lambda(t) = \frac{1}{2}t^2 + \log R(\hat{h}) - \hat{h}\sigma t = \frac{1}{6}\frac{\mu_3}{\sigma^3} t^3 + O(t^4), \tag{4.25}$$

where $|O(t^4)| \leqslant L(\gamma)t^4$. After substitution of (4.25) into (4.19), with $t = \sigma^{-1}n^{-1/2}y$ and $0 \leqslant y \leqslant y_n = o(n^{1/4})$, it is seen that

$$\exp\left\{n^{-1/2}\left(\frac{y}{\sigma}\right)^3\lambda\left(\frac{y}{\sigma n^{1/2}}\right)\right\} = \exp\left\{\frac{1}{6}n^{-1/2}\frac{\mu_3}{\sigma^3}\left(\frac{y}{\sigma}\right)^3 + O(n^{-1}y^4)\right\}$$

$$= \exp\left\{\frac{1}{6}n^{-1/2}\frac{\mu_3}{\sigma^3}\left(\frac{y}{\sigma}\right)^3\right\}\{1 + O(n^{-1}y^4)\},$$

where $|O(n^{-1}y^4)| \leqslant L(\gamma)n^{-1}y^4$. Thus all we need to do to complete the proof is show that the remainder term in (4.19) is uniformly of the stated order under the hypotheses of Proposition 4.1. Inspection of Petrov's (1995) Theorem 5.23 shows that the remainder term in (4.19) is exactly equal to

$$E_n(y) = (I_1 - I_3)/I_3 + (2\pi)^{1/2}I_2/I_3$$

where the quantities I_1, I_2 and I_3 are defined by Petrov. Under the assumption that $0 \leqslant t \leqslant \sigma\gamma/d_n$, it can be shown using arguments identical to those given by Petrov that

$$|E_n(y)| \leqslant L(\gamma)\hat{h}$$

where $\hat{h}$ is given in (4.24). Putting $t = \sigma^{-1}n^{-1/2}y$, where $0 \leqslant t \leqslant \sigma\gamma/d_n$, we find that for $0 \leqslant yd_n/\sigma^2 \leqslant n^{1/2}\gamma$, $|E_n(y)| = O\{n^{-1/2}(y+1)\}$ as required. □

Proof of Proposition 4.2. The proof is rather similar to that of Proposition 4.1. Using (C1′), again dropping the superscript (n), we have for fixed $(1 - A^{-1}) \leqslant \gamma < 1$, and $h \geqslant 0$, $hd_n \leqslant \gamma$,

$$R(h) = 1 + \frac{1}{2!}\sigma^2h^2 + \frac{1}{3!}\mu_3h^3 + \sum_{r=0}^{\infty}\frac{1}{(r+4)!}\mu_{r+4}h^{r+4}$$

$$= 1 + \frac{1}{2!}\sigma^2h^2 + \frac{1}{3!}\mu_3h^3 + O\left(h^4\sum_{r=0}^{\infty}d_n^rh^r\right)$$

$$= 1 + \frac{1}{2!}\sigma^2h^2 + \frac{1}{3!}\mu_3h^3 + O\{h^4(1-\gamma)^{-1}\}$$

$$= 1 + \frac{1}{2!}\sigma^2h^2 + \frac{1}{3!}\mu_3h^3 + O(h^4),$$

since γ is fixed and less than 1; and corresponding approximations hold for the derivatives of $R(h)$. From here on, the proof is almost identical to that of Proposition 4.1. □

Lemma 4.3. *Let F be an arbitrary distribution function. Then for any $\xi > 0$,*

$$I = \int_{-\infty}^{\infty} F(x-)^{\xi}\,\mathrm{d}F(x) \leqslant (1+\xi)^{-1}.$$

Proof. The result is true with equality when F is continuous, but this does not help with the general case. We shall establish the general case using an inductive argument. Consider an arbitrary two-point distribution $F = \pi_1 H_{x_1} + \pi_2 H_{x_2}$, where $\pi_j > 0$, $\pi_1 + \pi_2 = 1$, $x_1 < x_2$ and $H_x(y) = 0$ or 1 depending on whether $y \leqslant x$ or $y > x$. Then $I = \pi_1^\xi(1-\pi_1)$ has a maximum over $\pi_1 \in [0, 1]$ of $\{\xi/(1+\xi)\}^\xi(1+\xi)^{-1} \leqslant (1+\xi)^{-1}$ at $\pi_1 = \xi/(1+\xi)$. So the result holds for all two-point distributions. Suppose now that it holds for all m-point distributions $F = \sum_{j=1}^m \pi_j H_{x_j}$ where $\pi_j \geqslant 0$, $\sum \pi_j = 1$ and $x_1 \leqslant \ldots \leqslant x_m$. Then it holds for any $(m+1)$-point distribution $G = (1-\pi)F + \pi H_{x_{m+1}}$ with $\pi \geqslant 0$ and $x_{m+1} > x_m$, because

$$\begin{aligned} I &= (1-\pi)^\xi \pi + (1-\pi)^{\xi+1} \int_{-\infty}^{\infty} F(x-)^\xi \, \mathrm{d}F(x) \\ &\leqslant (1-\pi)^\xi \pi + (1-\pi)^{\xi+1}(1+\xi)^{-1}, \end{aligned}$$

using the induction assumption; and the last expression has a maximum over $\pi \in [0, 1]$ of $(1+\xi)^{-1}$ when $\pi = 0$. So the result holds for all finitely supported distributions. But, given an arbitrary distribution function F, we can find a sequence of distribution functions $(F^{(m)})$ such that $F^{(m)}$ has at most m points of support and also

$$\int F^{(m)}(x-)^\xi \, \mathrm{d}F^{(m)}(x) \to \int F(x-)^\xi \, \mathrm{d}F(x)$$

as $m \to \infty$. □

Lemma 4.4. *Let F be a distribution function and $r > 0$ a real number such that $\mu_r = \int_{-\infty}^{\infty} |x|^r \, \mathrm{d}F(x)$. If $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ are the order statistics of a random sample from F then, for any $0 < s < r$,*

$$U = \int_{(X_{(n)}, \infty)} |x|^s \, \mathrm{d}F(x) = O_p(n^{-(r-s)/r}).$$

Proof. It is sufficient to show that $\mathrm{E}(U) = O(n^{-(r-s)/r})$. Write $\beta = r/s$ and $\alpha = r/(r-s)$. Then using Fubini's theorem, Hölder's inequality and Lemma 4.3,

$$\begin{aligned} \mathrm{E}\int_{(X_{(n)}, \infty)} |x|^s \, \mathrm{d}F(x) &= \mathrm{E}\int_{-\infty}^{\infty} I(X_{(n)} < x)|x|^s \, \mathrm{d}F(x) \\ &= \int_{-\infty}^{\infty} F(x-)^n |x|^s \, \mathrm{d}F(x) \\ &\leqslant \left\{\int_{-\infty}^{\infty} F(x-)^{n\alpha} \, \mathrm{d}F(x)\right\}^{1/\alpha} \left\{\int_{-\infty}^{\infty} |x|^{s\beta} \, \mathrm{d}F(x)\right\}^{1/\beta} \\ &\leqslant (n\alpha+1)^{-1/\alpha}\{\mathrm{E}|X_1|^r\}^{1/\beta} \\ &= O(n^{-(r-s)/r}) \end{aligned}$$

as required. □

5. Proofs of theorems

Proof of Theorem 2.1. Write $d_n = \max\{|X_{(1)}|, |X_{(n)}|\}$. Since by hypothesis $\mathrm{E}X_1^4 < \infty$, Lemma 4.1 implies that $n^{-1/4} d_n \to 0$ in probability. Therefore we may apply Proposition 4.1 combined with Remark 4.1 twice, obtaining

$$\frac{\hat{G}_n(y)}{1 - \Phi(y/\sigma)} = \exp\left\{\frac{1}{6} n^{-1/2} \left(\frac{y}{\hat{\sigma}}\right)^3 \frac{\hat{\mu}_3}{\hat{\sigma}^3}\right\}\{1 + o_p(1)\} \tag{5.1}$$

and

$$\frac{H_n(y;\, X_{(1)},\, X_{(n)})}{1 - \Phi\{(y - n^{1/2}\tilde{\mu})/\tilde{\sigma}\}} = \exp\left\{\frac{1}{6} n^{-1/2} \left(\frac{y}{\tilde{\sigma}}\right)^3 \frac{\tilde{\mu}_3}{\tilde{\sigma}^3}\right\}\{1 + o_p(1)\}, \tag{5.2}$$

where, in both cases, the $o_p(1)$ remainder term is uniformly of the stated order for $0 \leqslant y \leqslant y_n$. In (5.1), $\hat{\sigma}^2$ and $\hat{\mu}_3$ are the variance and centred third moment of the sample $\mathscr{X}$; and in (5.2) $\tilde{\mu}$, $\tilde{\sigma}^2$ and $\tilde{\mu}_3$ are the theoretical mean, variance and centred third moment, respectively, of the truncated random variable $X_1 I(u \leqslant X_1 \leqslant v)$, evaluated at $u = X_{(1)}$ and $v = X_{(n)}$ after taking expectations. But $\mathrm{E}(X_1^4) < \infty$ implies that $\hat{\sigma}^2 = \sigma^2 + O_p(n^{-1/2})$ and $\hat{\mu}_3 = \mu_3 + O_p(n^{-1/4})$; and several applications of Lemma 4.4, using the assumptions $\mathrm{E}(X_1) = 0$ and $\mathrm{E}(X_1^4) < \infty$, imply that $\tilde{\mu} = O_p(n^{-3/4})$, $\tilde{\sigma}^2 = \sigma^2 + O_p(n^{-1/2})$ and $\tilde{\mu}_3 = \mu_3 + O_p(n^{-1/4})$. Consequently, since $0 \leqslant y \leqslant y_n = o(n^{1/4})$, $\hat{\sigma}^2$ and $\hat{\mu}_3$ in (5.1) may be replaced by σ^2 and μ_3, respectively, without changing the order of the error. Similarly, $\tilde{\mu}$, $\tilde{\sigma}^2$ and $\tilde{\mu}_3$ in (5.2) may be replaced by 0, σ^2 and μ_3 without changing the order of the error. The proof is now complete. □

Proof of Theorem 2.4. Let $b = b_n = -n^{1/4}$ and $a = a_n = \theta n^{1/2} y$, where $\theta \in (0, 1]$ will be specified in particular cases later. We first show that, with this choice of a and b, and with $y \geqslant 1$, $\Delta_F(b, a)\Delta_F(0, a)$ as $n \to \infty$. For $r \geqslant 1$,

$$\eta_{r+4}(b, 0) = \int_b^0 |x|^{r+4}\, \mathrm{d}F(x) \leqslant |b|^r \mathrm{E}(X_1^4)$$

and therefore

$$\Delta_F(b, 0) = \sup_{r \geqslant 1}\left\{\frac{\eta_{r+4}(b, 0)}{r!}\right\}^{1/r} \leqslant |b| \max\{1, \mathrm{E}(X_1^4)\}. \tag{5.3}$$

Since, for any $\varepsilon \in [0, 1]$ and $u, v \in \mathbb{R}$, the inequality $|u + v|^\epsilon \leqslant |u|^\epsilon + |v|^\epsilon$, it follows that for $r \geqslant 1$ and $y \geqslant 1$,

$$\eta_{r+4}(b, a)^{1/r} \leqslant \eta_{r+4}(b, 0)^{1/r} + \eta_{r+4}(0, a)^{1/r}$$

and therefore

$$\Delta_F(b, a) \leqslant \Delta_F(b, 0) + \Delta_F(0, a).$$

By (F2), g is slowly varying, so for any $\delta > 0$ and $y \geqslant 1$, $\Delta_F(0, a) \sim a/(g(a) \gg n^{(1/2)-\delta}$ and consequently, for such y,

$$\Delta_F(b, a) = \{1 + o(1)\}\Delta_F(0, a) \sim \Delta_F(0, a).$$

(i) Fix $\gamma \in (0, 1)$. If $0 \leqslant y \leqslant 1$ then the desired conclusion follows directly from the central limit theorem, so assume $y \geqslant 1$. In view of Lemma 4.2, part (i) will follow if we can show

$$P[n^{-1/2} S_n^{[b,a]} > y] \sim 1 - \Phi(y/\sigma) \tag{5.4}$$

and

$$P[X_{(n)} > a] = o\{1 - \Phi(y/\sigma)\} \tag{5.5}$$

uniformly for $1 \leqslant y \leqslant \gamma\lambda_n$.

To establish (5.4) we use Proposition (4.2). Choose $\theta = \frac{1}{2}$ in a and choose $d_n = \Delta_F(b, a)$ in (C1′). Since $d_n \sim a/g(a)$ and by assumption g is slowly varying, it follows that, given any $\varepsilon > 0$,

$$d_n \leqslant (1+\varepsilon)\frac{a}{g(a)} \quad \text{and} \quad \frac{g(n^{1/2}\lambda_n)}{g(\gamma n^{1/2}\lambda_n/2)} \leqslant 1 + \varepsilon$$

when n is sufficiently large. Choose ε so that $0 \leqslant \varepsilon < \gamma^{-1/2} - 1$. Then for $1 \leqslant y \leqslant \gamma\lambda_n$ and n sufficiently large,

$$\begin{aligned} \frac{n^{-1/2}y}{\sigma^2}\Delta_F(b, a) &\leqslant (1+\varepsilon)\frac{n^{-1/2}y}{\sigma^2}\frac{n^{1/2}y}{2g(n^{1/2}y/2)} \\ &\leqslant (1+\varepsilon)\gamma^2\frac{\lambda_n^2}{2\sigma^2 g(n^{1/2}\lambda_n)}\frac{g(n^{1/2}\lambda_n)}{g(\gamma n^{1/2}\lambda_n/2)} \\ &\leqslant (1+\varepsilon)^2\gamma^2 \leqslant \gamma < 1. \end{aligned} \tag{5.6}$$

Thus Proposition 4.2 may be applied and (5.4) is established.

To prove (5.5) note that for $y \geqslant 1$,

$$P[X_{(n)} > a] \sim n\{1 - F(a)\} \sim n\exp\{-g(a)\} \tag{5.7}$$

and

$$1 - \Phi(y/\sigma) \sim (2\pi)^{-1/2}\frac{\sigma}{y}\exp\left\{-\frac{y^2}{2\sigma^2}\right\}. \tag{5.8}$$

Choose $\theta = \frac{1}{2}$ in a. Following the kind of reasoning leading to (5.6), and using (F1)(c), it can be shown that

$$\sup_{1 \leqslant y \leqslant \gamma_n}\left\{\log y + y\frac{y^2}{2\sigma^2} + \log n - g(a)\right\} \to -\infty$$

which establishes (5.5).

(ii) Fix $\gamma > 1$. By Lemma 4.2,

$$P[n^{-1/2}S_n > y] \geqslant \{\tfrac{1}{2} + o(1)\}P[X_{(n)} > n^{1/2}y].$$

So, dividing (5.7) by (5.8) and noting that, by an argument similar to that leading to (5.6),

$$\frac{y^2}{2\sigma^2} - g(n^{1/2}y) = \frac{y^2}{2\sigma^2}\left(1 - \frac{2\sigma^2 g(n^{1/2}y)}{y^2}\right)$$
$$\geqslant (1-\gamma^{-1})\frac{y^2}{2\sigma^2},$$

part (ii) is proved. □

Proof of Theorem 2.5. The proof is quite similar to that of Theorem 2.4, but with some minor differences in detail. Fix $\theta = 1$ throughout the proof and recall that $\mathrm{E}(|X_1|^r) < \infty$ for some $r > \max\{4, 2(1-\alpha)^{-1}\}$, where α is given in condition (F3). Choose $b_n = b = -n^{1/r}$ and $a_n = a = n^{1/2}y$. Then following an argument similar to that used at the beginning of the proof of Theorem 2.4, it is seen that for $y \geqslant 1$,

$$\Delta_F(b, a) \sim \Delta_F(0, a) \sim \frac{a}{g(a)} \qquad \text{as } n \to \infty.$$

(When $0 \leqslant y < 1$, parts (i) and (iii) of the theorem are obviously true; and the condition $y \geqslant 1$ is not relevant to part (ii).) Also, note that under condition (F3),

$$\lambda_n \gg n^{\alpha/(4-2\alpha)-\varepsilon} \quad \text{and} \quad \lambda_n = o(n^{\alpha/(4-2\alpha)}) \tag{5.9}$$

for any fixed $\varepsilon > 0$.

(i) It follows from (5.9) that when $0 < \alpha < \frac{2}{3}$, $\lambda_n = o(y_n)$ where $y_n = o(n^{1/4})$ and $y_n \gg n^{(1/4)-\varepsilon}$ for all $\varepsilon > 0$. After applying Lemma 4.2 with $\theta = 1$, it is seen that part (i) will follow if

$$P[n^{-1/2}S_n^{[b,a]} > y] \sim 1 - \Phi(y/\sigma) \tag{5.10}$$

and

$$P[X_{(n)} > n^{1/2}y] = o\{1 - \Phi(y/\sigma)\} \tag{5.11}$$

uniformly for $1 \leqslant y \leqslant \gamma\lambda_n$, where

$$0 < \gamma < 2^{-1/(2-\alpha)}. \tag{5.12}$$

The proofs of (5.10) and (5.11) are very similar to those of (5.4) and (5.5) in Theorem 2.4(i); the only difference is that, instead of the requirement that $0 < \gamma < 1$, we need to use (5.12).

(ii) The proof is virtually identical to that of Theorem 2.4(ii).

(iii) When $\alpha > \frac{2}{3}$, $\lambda_n \gg y_n = o(n^{1/4})$. The proof is similar to that of Theorem 2.4(i) and Theorem 2.5(i). □

6. Some numerical results

The results of a small simulation study are now described. Three underlying distributions F were considered: $N(0, 1)$, labelled 'Nor'; and the lognormal $\exp\{\tau N(0, 1)\}$, labelled LN_1 and LN_2 for $\tau = 1, 2$, respectively. One hundred samples $\mathscr{X}_1, \ldots, \mathscr{X}_{100}$, each of size n, were generated from each of the three underlying distributions, for $n = 25$ and $n = 100$. Then Monte Carlo estimates of the quantities $\hat{\psi}_t$ and $\hat{\xi}_t$ $(t = 1, \ldots, 100)$ defined by

$$\hat{\psi}_t = P[n^{1/2}(\overline{X}_t^* - \overline{X}_t)/\sigma > 1.645|\mathscr{X}_t] \quad \text{and} \quad \hat{\xi}_t = P[n^{1/2}(\overline{X}_t^* - \overline{X}_t)/\hat{\sigma}_t^* > 1.645|\mathscr{X}_t]$$

were obtained, with each estimate based on 1000 bootstrap resamples. In the above, σ is the standard deviation of F; $\overline{X}_t$ is the mean of sample $\mathscr{X}_t$; and $\overline{X}_t^*$ and $\hat{\sigma}_t^*$ are the mean and standard deviation of a typical resample $\mathscr{X}_t^*$ obtained by random sampling, with replacement, from $\mathscr{X}_t$. The number 1.645 was chosen because $\Phi(1.645) = 0.95$, where Φ is the distribution function of $N(0, 1)$. The quantities $Q_1 \leqslant \ldots \leqslant Q_{100}$ are the ordered values of $\{\hat{\psi}_1, \ldots, \hat{\psi}_{100}\}$ in the percentile ('Per') cases, and the ordered values of $\{\hat{\xi}_1, \ldots, \hat{\xi}_{100}\}$ in the studentized ('Stu') cases. The final row of Table 1 contains Monte Carlo estimates of

$$\psi_0 = P[n^{1/2}(\overline{X} - \mu)/\sigma > 1.645] \quad \text{and} \quad \xi_0 = P[n^{1/2}(\overline{X} - \mu)/\hat{\sigma} > 1.645],$$

for each of the three underlying distributions, based on 10 000 simulated samples of size n from F in each case. The quantity $\hat{\sigma}$ in ξ_0 is the sample standard deviation.

Let us now discuss Table 1 in the percentile cases. Note that the bootstrap approximation to ψ_0 can be said to be performing 'well' if, columnwise, 'True' is close to Q_{50}, and Q_1 and Q_{100} are 'reasonably' close together; and the normal approximation to ψ_0 is performing well if ψ_0 is approximately 0.05. Bearing this in mind, the bootstrap approximation is

Table 1. Bootstrap estimates of some tail probabilities

	$n = 25$						$n = 100$					
	Nor		LN_1		LN_2		Nor		LN_1		LN_2	
	Per	Stu	Per	Stu	Per	Stu	Per	Stu	Per	Stu	Per	Stu
Q_1	0.001	0.033	0.000	0.006	0.000	0.002	0.020	0.039	0.000	0.005	0.000	0.003
Q_{25}	0.029	0.048	0.006	0.017	0.000	0.008	0.039	0.048	0.019	0.018	0.000	0.008
Q_{50}	0.043	0.057	0.019	0.025	0.000	0.013	0.049	0.052	0.040	0.025	0.000	0.014
Q_{75}	0.065	0.064	0.044	0.031	0.001	0.019	0.059	0.057	0.069	0.031	0.011	0.020
Q_{100}	0.115	0.111	0.173	0.056	0.228	0.033	0.095	0.071	0.262	0.046	0.282	0.035
True	0.052	0.057	0.063	0.011	0.025	0.001	0.050	0.051	0.057	0.018	0.037	0.002

Notes: 'Nor' indicates an underlying $N(0, 1)$ distribution, while 'LN_1' and 'LN_2' indicate $\exp\{\tau N(0, 1)\}$ distributions with $\tau = 1, 2$, respectively. The sample size is denoted by n. 'Per' denotes percentile, and 'Stu' denotes studentized. The quantities $Q_1 \leqslant \ldots \leqslant Q_{100}$ are the ordered values of 100 bootstrap estimates of the relevant tail probability (which has nominal value of 0.05), where each bootstrap estimate is based on 100 resamples; and 'True' denotes a direct Monte Carlo estimate of the true value of this tail probability, based on 10 000 simulated samples of size n.

performing well in the normal case, especially when $n = 100$; but performance is rather poor in the LN_1 cases and extremely poor in the LN_2 cases.

Recall the discussion of the lognormal case in Section 3, especially formula (3.1). Define $u(\tau, n) = (2\tau)^{-1} \log n$, and note that

$$u(1, 25) = 1.61, \qquad u(1, 100) = 2.30, \qquad u(2, 25) = 0.80 \quad \text{and} \quad u(2, 100) = 1.15.$$

Comparing 1.645 with $u(\tau, n)$ in the cases considered, we see that relative error considerations do anticipate the poor bootstrap performance in the lognormal cases.

Nevertheless, some care is needed, and a little more needs to be said. The poor performance is not due to the breakdown of the normal approximation to ψ_0, as in each case ψ_0 is fairly close to the nominal value of 0.05 (cf. Theorem 2.4 and Corollary 2.2). Further numerical results (which are not given here) show that the bootstrap approximation is actually quite good when $\hat{\sigma}$ rather than σ is used in the calculation of the $\hat{\psi}_t$, and that the problem is that $\hat{\sigma}/\sigma$ is typically quite different from 1 in the lognormal cases considered. Since the asymptotic results in Section 2 do not really distinguish between $\hat{\sigma}$ and σ to the order of error considered, this is, in a sense, a 'subasymptotic' difficulty.

One might speculate that this scaling problem can be dealt with by studentizing. The effects of studentizing are investigated in the columns of Table 1 labelled 'Stu'. It can be seen that, in the normal cases, studentizing has a beneficial effect. However, in the LN_1 cases studentizing 'over-compensates', so that there is no overall improvement; and in the LN_2 cases studentizing actually makes matters substantially worse.

7. Discussion

In this paper we have presented results on the relative error properties of the bootstrap approximation to the distribution of the sample mean when the underlying distribution is sub-exponential. In particular, we have determined a sequence (λ_n), depending on the underlying distribution in a simple way, which determines at what point in the tail the bootstrap approximation breaks down. Although the breakdown does not appear to be as sharp in practice as the theory suggests, the results of a small simulation study show that the sequence λ_n does have practical relevance (even though λ_n is obtained via asymptotic considerations).

The problem of developing useful diagnostics for predicting bootstrap performance is one which deserves further study. It is hoped that more detailed study of the errors in (2.5), Lemma 4.2 and Proposition 4.2 might ultimately lead to useful diagnostics of this type, at least in simpler settings.

On the theoretical side, it would be nice to know more about the relative error and breakdown properties in the case of statistics more complicated than a mean, e.g. a studentized mean, and to extend any such results to the relative coverage error of the various types of bootstrap confidence interval.

References

Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation*. Cambridge: Cambridge University Press.

Booth, J.G., Hall, P. and Wood, A.T.A. (1994) On the validity of Edgeworth and saddlepoint approximations. *J. Multivariate Anal*., **51**, 121–138.

Efron, B. (1979) Bootstrap methods: another look at the jacknife. *Ann. Statist*., **7**, 1–26.

Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Hahn, M.G. and Klass, M.J. (1997) Approximation of partial sums of arbitrary I.I.D. random variables and the precision of the usual exponential upper bound. *Ann. Probab*., **25**, 1451–1470.

Hall, P. (1986) On the bootstrap and confidence intervals. *Ann. Statist*., **14**, 1431–1452.

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

Hall, P. (1990) On the relative performance of bootstrap and Edgeworth approximations of a distribution function. *J. Multivariate Anal*., **35**, 108–129.

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

Jing, B.Y. (1992) Saddlepoint and Edgeworth approximations and their applications to resampling. Unpublished Ph.D. thesis, University of Sydney.

Jing, B.Y., Feuerverger, A. and Robinson, J. (1994) On the bootstrap saddlepoint approximation. *Biometrika*, **81**, 211–215.

Lee, S.M.S. and Young, G.A. (1995) Asymptotic iterated bootstrap confidence intervals. *Ann. Statist*., **23**, 1301–1330.

Nagaev, A.V. (1969) Limit theorems for large deviations when Cramér's conditions are violated (in Russian). *Izv. Akad. Nauk. UzSSR Ser. Fiz.-Mat. Nauk*., **6**, 17–22.

Nagaev, S.V. (1973) Large deviations for sums of independent random variables. In J. Koževšník (ed.), *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, pp. 657–674, Prague: Academia.

Nagaev, S.V. (1979) Large deviations of sums of independent random variables. *Ann. Probab*., **7**, 745–789.

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

Petrov, V.V. (1995) *Limit Theorems of Probability Theory*, Oxford Studies in Probability 4. Oxford: Clarendon Press.

Pitman, E.J.G. (1980) Subexponential distribution functions. *J. Austral. Math. Soc. Ser. A*, **29**, 337–347.

Singh, K. (1981) On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist*., **9**, 1189–1195.

Teugels, J.L. (1975) The class of subexponential distributions. *Ann. Probab*., **3**, 1000–1011.

Received February 1998 and revised June 1999