Metrika (1999) 50: 55–69

Metrika

Subsampling inference for the mean in the heavy-tailed case*

Joseph P. Romano[1], Michael Wolf[2]

[1] Department of Statistics, Stanford University, Sequoia Hall, Stanford, CA 94305, USA
(e-mail: romano@stat.stanford.edu)
[2] Departamento de Estadística y Econometría, Universidad Carlos III de Madrid,
Calle Madrid 126, E-28903 Getafe, Spain (e-mail: mwolf@est-econ.uc3m.es)

Received: December 1998

Abstract. In this article, asymptotic inference for the mean of i.i.d. observations in the context of heavy-tailed distributions is discussed. While both the standard asymptotic method based on the normal approximation and Efron's bootstrap are inconsistent when the underlying distribution does not possess a second moment, we propose two approaches based on the subsampling idea of Politis and Romano (1994) which will give correct answers. The first approach uses the fact that the sample mean, properly standardized, will under some regularity conditions have a limiting stable distribution. The second approach consists of subsampling the usual t-statistic and is somewhat more general. A simulation study compares the small sample performance of the two methods.

Key words: Heavy tails, self-normalization, stable laws, subsampling

1 Introduction

It has been two decades since Efron (1979) introduced the bootstrap procedure for estimating sampling distributions of statistics based on independent and identically distributed (i.i.d.) observations. While the bootstrap has enjoyed tremendous success and has led to something like a revolution of the field of statistics, it is known to fail for a number of counterexamples. One well-known example is the case of the mean when the observations are heavy-tailed. If the observations are i.i.d. according to a distribution in the domain of attraction of a stable law with index $\alpha < 2$ (see Feller, 1971), then the sample mean appropriately normalized converges to a stable law. However, Athreya (1987) showed that the bootstrap version of the normalized mean has

* The final version of this paper has benefited from helpful comments of two anonymous referees.

a limiting random distribution, implying inconsistency of the bootstrap. An alternative proof of Athreya's result was presented by Knight (1989). Kinateder (1992) gave an invariance principle for symmetric heavy-tailed observations. It has been realized that taking a smaller bootstrap sample size can result in consistency of the bootstrap; but knowledge of the tail index of the limiting law is needed. See Athreya (1985) and Arcones (1990); also see Wu, Carlstein, and Cambanis (1993) and Arcones and Giné (1989).

In this article, we describe how the subsampling method (Politis and Romano, 1994) can be used to make asymptotically correct inference for the mean in the heavy-tailed case without knowledge of the tail index. We present two different approaches. The first one appeals to the limiting stable distribution when the sample mean is normalized accordingly. This involves knowledge or estimation of the tail index of the underlying distribution. The second approach uses the idea of self-normalizing sums (e.g., Logan et al., 1973), avoiding the explicit estimation of the tail index.

The paper is organized as follows. In Section 2, we present an extension of the general subsampling theory which allows to subsample studentized statistics when the scale estimator does not converge in probability; this extension is needed for the approach utilizing self-normalizing sums. The two explicit subsampling approaches for making inference for the univariate mean in the context of heavy-tailed observations are discussed in Section 3. We propose a method for choosing the block size in Section 4. A simulation study in Section 5 sheds some light on small sample performance. Conclusions are stated in Section 6. All tables and figures appear at the end of the paper.

2 The subsampling method

2.1 Standard theory

The subsampling methodology was introduced by Politis and Romano (1994) as an inference procedure that allows to construct asymptotically valid confidence regions under very weak assumptions. We will briefly describe the basic method before presenting an extension of the general theory pertaining to studentized statistics.

Consider a random sample of i.i.d. variables $X_1, \dots, X_n$ in an arbitrary sample space S. Denote the common underlying probability measure by P. The goal is to construct a confidence interval for some parameter $\theta = \theta(P) \in \mathbb{R}$. Let $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ be an estimator of θ. No assumptions on the form of the estimator are made, although it seems natural in the i.i.d. context to use an estimator that is symmetric in its arguments.

The basic subsampling method consists of approximating the sampling distribution of $\hat{\theta}_n - \theta$ by computing the estimator on smaller subsets (or subsamples) of the observed data and using the empirical distribution of these subsample values after an appropriate normalization. To be more specific, for an integer $b < n$, let $Y_1, \dots, Y_{N_{n,b}}$ be equal to the $N_{n,b} = \binom{n}{b}$ subsets of size b of $\{X_1, \dots, X_n\}$, ordered in any fashion. Now, let $\hat{\theta}_{b,i}$ be equal to the statistic $\hat{\theta}_b$ evaluated at the data set Y_i. Then, the subsampling approximation of

$$Prob_P\{\tau_n(\hat{\theta}_n - \theta) \le x\}$$

is given by

$$L_{n,b}(x) = N_{n,b}^{-1} \sum_{i=1}^{N_n} 1\{\tau_b(\hat{\theta}_{b,i} - \hat{\theta}_n) \le x\}. \tag{1}$$

Here, τ_n and τ_b are appropriate normalizing constants chosen such that $\tau_n(\hat{\theta}_n - \theta)$ has a nondegenerate limiting distribution. Hence, τ_n is the rate of convergence of $\hat{\theta}_n$ and in regular cases we have $\tau_n = n^{1/2}$.

The quantiles of $L_{n,b}(\cdot)$ can then be used to construct approximate one-sided confidence intervals for θ in the obvious fashion. These intervals will have asymptotically correct coverage given that $\tau_n(\hat{\theta}_n - \theta)$ has a non-degenerate limiting distribution and that $\tau_n/\tau_n \to 0$, $b/n \to 0$, and $b \to \infty$ as $n \to \infty$; see Politis and Romano (1994) for details.

When two-sided confidence intervals for θ are desired, they can be constructed by the intersection of two one-sided intervals, resulting in so-called equal-tailed intervals. An alternative procedure is to construct symmetric intervals – extending equally far to the left and to the right of the point estimate $\hat{\theta}_n$ – by estimating the two-sided sampling distribution function

$$Prob_P\{\tau_n|\hat{\theta}_n - \theta| \le x\}.$$

The corresponding subsampling approximation is then given by

$$L_{n,b,|\cdot|}(x) = N_{n,b}^{-1} \sum_{i=1}^{N_n} 1\{\tau_b|\hat{\theta}_{b,i} - \hat{\theta}_n| \le x\}. \tag{2}$$

Given the existence of an Edgeworth expansion, symmetric subsampling intervals often exhibit improved coverage properties; for example, see Chapter 10 of Politis, Romano, and Wolf (1999).

Note that the exact calculations of $L_{n,b}(x)$ and $L_{n,b,|\cdot|}(x)$ are prohibitive for moderate or large sample sizes, since they require the evaluation of $N_{n,b} = \binom{n}{b}$ subsample statistics. However, a stochastic approximation may be used instead. Let $I_1, \dots, I_s$ be chosen randomly with or without replacement from $\{1, 2, \dots, N\}$. Then, $L_{n,b}(x)$ may be approximated by

$$\hat{L}_{n,b}(x) = s^{-1} \sum_{i=1}^{s} 1\{\tau_b(\hat{\theta}_{b,I_i} - \hat{\theta}_n) \le x\} \tag{3}$$

and $L_{n,b,|\cdot|}(x)$ may be approximated analogously. These stochastic approximations do not affect the asymptotic validity of the method provided that $s \to \infty$ as $n \to \infty$.

The application of the basic subsampling method requires knowledge of the rate of convergence τ_n. But, for our application of the mean of heavy-tailed observations, it is well known that the rate of convergence depends on the tail index of the limiting law and is therefore unknown in practice; see Proposition 3.1. One way out of this dilemma is to estimate the rate of convergence from the data and use the estimated rate in the construction of the subsampling distribution. This is the general idea of Bertail, Politis, and Romano (1999); see Subsection 3.1. Another solution is to consider the usual t-statistic for the sample mean which turns out the a self-normalized sum, that is, it always has a proper limiting distribution. Subsampling studentized statistics when the estimate of scale converges in probability was considered by Politis and Romano (1993). This covers the situation of i.i.d. data with finite variance but not the heavy-tailed case. The following subsection provides an extension of the theory that allows for the estimate of scale to converge in distribution.

2.2 Subsampling studentized statistics

Focus is now on a studentized statistic $\tau_n^*(\hat{\theta}_n - \theta)/\hat{\sigma}_n$, where $\hat{\sigma}_n$ is some positive estimate of scale. Note that the appropriate normalizing constant τ_n^* may be different from its counterpart τ_n in the non-studentized case. Define $J_n^*(P)$ to be the sampling distribution of $\tau_n^*(\hat{\theta}_n - \theta)/\hat{\sigma}_n$ based on a sample of size n from P. Also define the corresponding cumulative distribution function

$$J_n^*(x, P) = Prob_P\{\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n \leq x\}.$$

The essential assumption needed to construct asymptotically valid confidence regions for θ now becomes slightly more involved than for the non-studentized case.

Assumption 2.1. *$J_n^*(P)$ converges weakly to a limit law $J^*(P)$. In addition, $a_n(\hat{\theta}_n - \theta(P))$ converges weakly to V, and $d_n\hat{\sigma}_n$ converges weakly to W, for positive sequences $\{a_n\}$ and $\{d_n\}$ satisfying $\tau_n = a_n/d_n$. Here, V and W are two random variables, where W does not have positive mass at zero.*

The subsampling method is modified to the studentized case in the obvious way. Let $\hat{\sigma}_{b,i}$ be equal to the estimate of scale based on the subsample Y_i. Analogous to (1) define

$$L_{n,b}^*(x) = N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{\tau_b(\hat{\theta}_{b,i} - \hat{\theta}_n)/\hat{\sigma}_{b,i} \leq x\}. \tag{4}$$

$L_{n,b}^*(x)$ then represents the subsampling approximation to $J_n^*(x)$. The following theorem shows that this approximation leads to asymptotically correct confidence intervals for θ.

Theorem 2.1. *Assume Assumption 2.1, $a_b/a_n \to 0$, $\tau_b/\tau_n \to 0$, $b/n \to 0$ and $b \to \infty$ as $n \to \infty$. Let x be a continuity point of $J^*(\cdot, P)$. Then*

(i) $L_{n,b}^*(x) \to J^*(x, P)$ in probability.
(ii) If $J^*(\cdot, P)$ is continuous, then $\sup_x |L_{n,b}^*(x) - J^*(x, P)| \to 0$ in probability.
(iii) For $\alpha \in (0, 1)$, let $c_{n,b}(1 - \alpha) = \inf\{x : L_{n,b}^*(x) \geq 1 - \alpha\}$. Correspondingly, define $c(1 - \alpha, P) = \inf\{x : J^*(x, P) \geq 1 - \alpha\}$. If $J^*(\cdot, P)$ is con-

tinuous at $c(1-\alpha, P)$ *then* $Prob_P\{\tau_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n \leq c_n(1-\alpha)\} \to 1-\alpha$ *as* $n \to \infty$. *Thus, the asymptotic coverage probability under P of the interval* $I_1 = [\hat{\theta}_n - \hat{\sigma}_n \tau_n^{-1} c_n(1-\alpha), \infty)$ *is the nominal level* $1-\alpha$.

Proof: To prove (i), note that

$$L^*_{n,b}(x) = N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{\tau_b(\hat{\theta}_{n,i} - \hat{\theta}_n)/\hat{\sigma}_{n,i} \leq x\}$$

$$= N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{\tau_b(\hat{\theta}_{n,i} - \theta)/\hat{\sigma}_{n,i} \leq x + \tau_b(\hat{\theta}_n - \theta)/\hat{\sigma}_{n,i}\}. \tag{5}$$

We want to show that the terms $\tau_b(\hat{\theta}_n - \theta)/\hat{\sigma}_{n,i}$ are negligible in the last equation. To this end, for $t > 0$, let

$$R_{n,b}(t) = N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{\tau_b(\hat{\theta}_n - \theta)/\hat{\sigma}_{n,i} \leq t\}$$

$$= N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{d_b\hat{\sigma}_{n,i} \geq d_b\tau_b(\hat{\theta}_n - \theta)/t\}$$

$$= N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{d_b\hat{\sigma}_{n,i} \geq a_b(\hat{\theta}_n - \theta)/t\}.$$

Here, we are making use of the fact that both the sequences a_n and b_n are positive. By Assumption 2.1 and $a_b/a_n \to 0$, we have for any $\delta > 0$ that $a_b(\hat{\theta}_n - \theta) \leq \delta$ with probability tending to one. Therefore, with probability tending to one

$$R_{n,b}(t) \geq N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{d_b\hat{\sigma}_{n,i} \geq \delta/t\}.$$

We need to consider the case $t > 0$ only, as the scale estimates $\hat{\sigma}_{n,i}$ are positive. Due to the usual subsampling argument (Politis and Romano, 1994, Theorem 2.1), $N_{n,b}^{-1}\sum_{i=1}^{N_{n,b}} 1\{d_b\hat{\sigma}_{n,i} \geq \delta/t\}$ converges in probability to $P(W \geq \delta/t)$, as long as δ/t is a continuity point of W. Hence, we can make sure that $R_{n,b}(t)$ is arbitrarily close to one by choosing δ small enough; remember we assume that W does not have positive mass at zero. In other words, for any $t > 0$, we have $R_{n,b}(t) \to 1$ in probability. Let us now rewrite (5) in the following way

$$L^*_{n,b}(x) = N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{\tau_b(\hat{\theta}_{n,i} - \theta)/\hat{\sigma}_{n,i} \leq x + \tau_b(\hat{\theta}_n - \theta)/\hat{\sigma}_{n,i}\}$$

$$\leq N_{n,b}^{-1} \sum_{i=1}^{N_{n,b}} 1\{\tau_b(\hat{\theta}_{n,i} - \theta)/\hat{\sigma}_{n,i} \leq x + t\} + (1 - R_{n,b}(t)),$$

for any positive number t. The last inequality follows because the i-th term in (5) is less than or equal to

$$1\{\tau_b(\hat{\theta}_{n,i} - \theta)/\hat{\sigma}_{n,i} \le x + t\} + 1\{\tau_b(\hat{\theta}_n - \theta)/\hat{\sigma}_{n,i} > t\}; \tag{6}$$

then, sum over all i. We have seen that $(1 - R_{n,b}(t)) \to 0$ in probability and hence by a standard subsampling argument again we get, for any $\varepsilon > 0$, $L^*_{n,b}(x) \le J^*(x + t, P) + \varepsilon$ with probability tending to one, provided that $x + t$ is a continuity point of $J^*(\cdot\,, P)$. Letting t tend to zero shows that $L^*_{n,b}(x) \le J^*(x, P) + \varepsilon$ with probability tending to one. A similar argument leads to $L^*_{n,b}(x) \ge J^*(x, P) - \varepsilon$ with probability tending to one. Since ε is arbitrary, this implies $L^*_{n,b}(x) \to J^*(x, P)$ in probability, and thus we have proved (i).

The proofs of (ii) and (iii) given (i) are very similar to the proofs of (ii) and (iii) given (i) in Theorem 2.1 of Politis and Romano (1994) and thus are omitted. ■

Remark 2.1. The issues of symmetric confidence intervals and stochastic approximation, as discussed for the case of the non-studentized subsampling method, apply as well and the corresponding results are analogous.

3 Subsampling inference for the mean

Suppose the X_i are i.i.d. univariate random variables in the domain of attraction of a stable law with index $1 < \alpha \le 2$. For a detailed discussion of stable distributions, the reader is referred to Zolotarev (1986) and Samorodnitsky and Taqqu (1994). When $1 < \alpha < 2$, it follows that the underlying distribution P possesses a finite mean but that its variance is infinite. The goal is to find a confidence interval for $\theta = E(X_i)$. Our choice for the estimator is the sample mean $\hat{\theta}_n = \overline{X}_n = n^{-1} \sum_{i=1}^n X_i$. The subsampling methodology requires a normalization resulting in a nondegenerate limiting distribution. In this section, we will discuss two possible approaches, one that relies on a stable limiting law and another one which uses self-normalizing sums.

3.1 Appealing to a limiting stable law

In case the underlying distribution belongs to the normal domain of attraction of a stable law, we can make use of the following result.

Proposition 3.1. *Assume $X_1, X_2, \ldots$ is a sequence of random variables in the normal domain of attraction of a stable law with index of stability $1 < \alpha \le 2$. Denote the common mean by θ. Then, $n^{-1/\alpha}(X_1 + \cdots + X_n - n\theta) = n^{1-1/\alpha}(\overline{X}_n - \theta)$ converges weakly to an α-stable distribution with mean zero.*

Proof: The proof follows immediately from the CLT when $\alpha = 2$. For the case of $1 < \alpha < 2$, it is a consequence of Theorem 3 in section XVII.5 of Feller (1971). ■

One might be tempted to use this result to construct confidence intervals for θ based on the quantiles of the (estimated) limiting stable distribution.

However, in addition to the stable index α, this distribution also depends on a skewness and on a scale parameter which are very difficult to estimate.

On the other hand, the non-studentized subsampling technique only requires knowledge or a consistent estimate of the index α, since the normalizing constants are given by $\tau_n = n^{1-1/\alpha}$ and $\tau_b = b^{1-1/\alpha}$, respectively. Let $\hat{\alpha}_n = \hat{\alpha}_n(X_1, \ldots, X_n)$ denote an estimator of α based on the segment $X_1, \ldots, X_n$. This notation includes the (rare) case where α is known, since $\hat{\alpha}_n \equiv \alpha$ is a valid estimator. Then the subsampling approximation of

$$Prob_P\{\tau_n(\overline{X}_n - \theta) \leq x\}$$

is given by

$$L_n^{\hat{\alpha}}(x) = N_n^{-1} \sum_{i=1}^{N_n} 1\{b^{1-1/\hat{\alpha}_n}(\overline{X}_{b,i} - \overline{X}_n) \leq x\}, \tag{7}$$

where $\overline{X}_{b,i} = b^{-1} \sum_{j=i}^{i+b-1} X_j$. Given that $\hat{\alpha}_n = \alpha + o_P((\log n)^{-1})$, this approximation can be used to construct asymptotically valid confidence intervals for θ; see Theorem 5 of Bertail et al. (1999).

Therefore, applying the subsampling method only requires a $\log n$ consistent estimator for the tail index α. Several such estimators are known, among them the Pickands (1975), Hill (1975), and deHaan and Resnick (1980) estimators. Tail index estimators typically are based upon a number q of extreme order statistics. Asymptotic consistency of the estimators requires that $q \to \infty$ but $q/n \to 0$ as $n \to \infty$. Unfortunately, the choice of q in practice is a very difficult problem and its effect can be tremendous even for sample sizes above thousand; for example, see Mittnik et al. (1996) and Resnick (1997).

At this point, we propose an alternative tail index estimator based on the subsampling technique. As noticed before, when the underlying distribution is in the normal domain of attraction of a stable law, the proper normalizing constant is $n^{1-1/\alpha}$ so that the rate of convergence is $\beta \equiv 1 - 1/\alpha$. Bertail et al. (1999) discuss consistent subsampling estimators for the rate of convergence of general statistics. These estimators depend on a number I of subsampling distributions with different block sizes, where $I \geq 2$, and on a number J of corresponding estimated quantiles, where $J \geq 1$. In the paper basically two estimators are described, the Quantile estimator and the Range estimator. We will use the Range estimator, denoted by $\hat{\beta}_{I,j}$. See Bertail et al. (1999) for the derivation of this estimator and a proof of its consistency. In their Theorem 2 it is shown that under mild regularity conditions, $\beta_{I,J} = \beta + o_P((\log n)^{-1})$.

For our application of the mean in the heavy-tailed context, it was seen that $\beta = 1 - 1/\alpha$, as long as the underlying distribution is in the normal domain of attraction of a stable law with tail index α. Hence, an obvious estimator of α is given as

$$\hat{\alpha}_{I,J} = 1/(1 - \hat{\beta}_{I,J}). \tag{8}$$

It immediately follows that under the same regularity conditions of Theorem 2 of Bertail et al. (1999), we have $\hat{\alpha}_{I,J} = \alpha + o_P((\log n)^{-1})$.

Remark 3.1. As an alternative to subsampling, the bootstrap with resample size $m < n$ may be considered. However, as in the case of subsampling, the proper standardization of the bootstrap distribution depends on the underlying tail index. In the case where the standardization is known, Athreya (1985) showed that the bootstrap distribution converges to the right limit in probability, given that $m \to \infty$ and $m/n \to 0$. Arcones and Giné (1989) strengthened this result to almost sure convergence under the condition of $m(\log\log n)/n \to 0$. Neither paper discusses the validity of the bootstrap approach in conjunction with an estimated rate. Also, no suggestion of how to pick the resample size m in practice is made. A general discussion of the bootstrap with resample size $m < n$ can be found in Bickel, Götze, and van Zwet (1997).

Wu, Carlstein, and Cambanis (1993) introduced an averaged bootstrap that overcomes the randomness in the limiting law of the bootstrap with resample size $m = n$. They showed that the averaged bootstrap converges to the correct limit in the case of heavy-tailed data having an exact stable distribution, provided that appropriate sample-based adjustments for scale and skewness are made. However, this approach may not extend to distributions in the domain of attraction of stable laws.

3.2 Using self-normalizing sums

It is well known that if the observations are i.i.d. from a distribution with finite second moment, then the t-statistic

$$T_n = n^{1/2}\frac{\overline{X}_n - \theta}{S_n}$$

has a limiting standard normal distribution. Here, S_n is the square root of the usual estimate of variance

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

The fact that T_n has, under fairly general conditions, a nondegenerate limiting distribution even if the underlying distribution has an infinite second moment makes it a *self-normalizing sum*. The limiting behavior of T_n for heavy-tailed distributions has, among others, been studied by Hotelling (1961), Efron (1969), and Logan et al. (1973). In the paper of Logan et al. (1973), exact densities of the limiting distribution of T_n are derived for the case of the underlying distribution belonging to the domain of attraction of a stable law. It is seen that the density does not only depend in a complicated way on the tail index α but also on some other characteristics of the limiting distribution. Again, this greatly diminishes the appeal of any inference based on the explicit estimation of the limiting distribution. On the other hand, the following proposition allows for an easy application of the subsampling method.

Proposition 3.2. *Assume* $\{X_i\}$ *is a sequence of i.i.d. random variables in the domain of attraction of an* α*-stable law with* $1 < \alpha \leq 2$. *Denote the common* *mean by θ. Define $\hat{\theta}_n = \bar{X}_n$, the usual sample mean, and $\hat{\sigma}_n = S_n$, the usual sample standard deviation. Also, let $\tau_n = n^{1/2}$ and $\tau_b = b^{1/2}$. Assume that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$.*

Then the conclusions of Theorem 2.1 hold.

Proof: We have to show that the conditions of Assumption 2.1 are met. To this end define

$$U_n = n^{1/2} \frac{\bar{X}_n - \theta}{(n^{-1} \sum_{i=1}^{n} (X_i - \theta)^2)^{1/2}} \tag{9}$$

$$= \frac{V_n}{W_n}, \tag{10}$$

where

$$V_n = \frac{X_1 + \cdots + X_n - n\theta}{n^{1/\alpha} L(n)} \quad \text{and} \quad W_n = \left(\frac{(X_1 - \theta)^2 + \cdots + (X_n - \theta)^2}{n^{2/\alpha} L^2(n)} \right)^{1/2}. \tag{11}$$

Here, $L(\cdot)$ is a slowly varying function ensuring that V_n converges to a stable law G; for example, see Feller (1971, Section XVII.5).

First consider the case where the X_i have a stable distribution. Logan et al. (1973) show that in this case (V_n, W_n) has a nondegenerate joint limiting distribution, where the limiting distribution of W_n does not have positive mass at zero. Indeed, the limiting distribution of W_n^2 is a positive stable law with index $\alpha/2$.

It turns out that in the general case, where the X_i are in the domain of attraction of G, the joint limiting distribution of (V_n, W_n) is identical to that of the stable case. Again, see Logan et al. (1973).

By simple algebra, finally

$$T_n = U_n \left(\frac{n-1}{n - W_n^2} \right)^{1/2},$$

where the second term converges to one in probability. Hence the conditions of Assumption 2.1 are satisfied. ■

The power of Proposition 3.2 lies in the fact that we always can subsample the t-statistic T_n, regardless of the tail index α of the underlying distribution. Therefore, it is not necessary to know or to estimate α. In addition this approach is not restricted to distributions in the *normal* domain of attraction of stable laws and therefore it is more general than the method of Subsection 3.1.

Remark 3.2. Arcones and Giné (1991) showed that bootstrapping the t-statistic with resampling size $m < n$ will also give asymptotically correct results, given that $m \to \infty$ and $m/n \to 0$. They suggested to choose $m = n/(\log \log n)^{1+\delta}$ for some small $\delta > 0$.

4 Choice of the block size

A practical issue in using the subsampling method is the choice of the block size b. Politis, Romano, and Wolf (1997) propose a calibration technique that corrects for over- or undercoverage of subsampling intervals for finite samples by adjusting the nominal confidence level accordingly. This technique involves generating pseudo sequences $X_1^*, \ldots, X_n^*$ using a suitable bootstrap method – Efron's (1979) bootstrap for i.i.d. data or Künsch's (1989) moving blocks bootstrap for time series data. Hence, this idea is limited to applications where standard bootstrap methods are also consistent and would fail for our problem at hand. We therefore present a different method which will work under more general conditions, that is, whenever subsampling applies.

Our method is of heuristic nature and we do not claim any optimality properties. It is based on the fact that for the subsampling method to be consistent the block size b needs to tend to infinity with the sample size n but a smaller rate, satisfying $b/n \to 0$. For b too close to n all subsample statistics $\hat{\theta}_{n,i}$ will be almost equal to $\hat{\theta}_n$, resulting in the subsampling distributions $L_{n,b}$ or $L_{n,b}^*$ being too tight and in undercoverage of subsampling confidence intervals. Indeed, for very large block sizes the confidence intervals will shrink towards the singleton $\hat{\theta}_n$, which consequence of the fact that the subsampling distributions $L_{n,b}(\cdot)$ and $L_{n,b}^*(\cdot)$ both collapse to a point mass at zero as the block size b tends to n (e.g., Lahiri, 1998). On the other hand, if b is too small, the intervals can undercover or overcover depending on the state of nature (e.g., Politis, Romano, and Wolf, 1997). This leaves a number of b values in the "right range" where we would expect almost correct results, at least for big sample sizes. We exploit this idea by computing subsampling intervals for a large number of block sizes b and then looking for a region where the intervals do not change very much. Within this region we then pick one interval according to some arbitrary criterion.

While this method can be carried out by "visual inspection" it is desirable to also have some automatic selection procedure, at the very least when simulation studies are to be carried out. The procedure we propose is based on minimizing a running standard deviation. Assume we compute subsampling intervals for block sizes b in the range of b_{small} to b_{big}. The endpoints of the confidence intervals should change in smooth fashion, as b changes. This might be somewhat violated if we use a stochastic approximation, such as (3), for moderate or large sample sizes. In that case it seems sensible to enforce some smoothness by applying a running mean to the endpoints of the intervals. A running standard deviation applied to the endpoints then determines the volatility around a specific b value. We choose the value of b with the smallest volatility. Here is a more formal description of the algorithm.

Algorithm 4.1 (Minimizing Confidence Interval Volatility)

1. For $b = b_{small}$ to $b = b_{big}$ compute a subsampling interval for θ at the desired confidence level, resulting in endpoints $I_{b,low}$ and $I_{b,up}$.
2. If a stochastic approximation such as (3) was used in Step 1, smooth the lower and upper endpoints separately, using a running mean of span m. This means replace $I_{b,low}$ by the average of $\{I_{b-m,low}, I_{b-m+1,low}, \ldots, I_{b+m,low}\}$ and do the same for $I_{b,up}$.

3. For each b compute a volatility index VI_b as the standard deviation of the interval endpoints in a neighborhood of b. More specifically, for a small integer k, let VI_b be equal to the standard deviation of $\{I_{b-k,low}, \ldots, I_{b+k,low}\}$ plus the standard deviation of $\{I_{b-k,up}, \ldots, I_{b+k,up}\}$.
4. Pick the value b^* with the smallest volatility index and report $[I_{b^*,low}, I_{b^*,up}]$ as the final confidence interval.

Some remarks concerning the implementation of this algorithm are in order.

Remark 4.1. The range of b values, determined by b_{small} and b_{big}, which is included in the minimization algorithm is not very important, as long as it is not too narrow.

Remark 4.2. To make the algorithm more computationally efficient, it might be desirable to skip a number of b values in a regular fashion. For example, include only every other b between b_{small} and b_{big}.

Remark 4.3. The algorithm contains two model parameters, m and k. Simulation studies have shown that the algorithm is very insensitive to both parameters. We usually employ $m = 2$ or $m = 3$ and the same for k.

We now illustrate how the algorithm works with the help of two simulated data sets. First, we generated a data set of size $n = 100$ i.i.d. from a symmetric stable distribution with mean zero and tail index $\alpha = 1.5$. The range of b values was chosen as $b_{small} = 4$ and $b_{big} = 40$. We computed symmetric subsampling intervals according to the approaches of Subsection 3.1, taking α to be known, and of Subsection 3.2, avoiding the knowledge of explicit estimation of α. Since the stochastic approximation of the kind (3) was employed with $N = 1000$, we smoothed the endpoints according to Step 2 and $m = 2$. The minimization of the volatility in Step 3 was done using $k = 2$. The results are shown at the top of Figure 1. The left plot corresponds to the approach appealing to a limiting stable law, while the right plot corresponds to the self-normalizing approach. The block sizes b chosen by the algorithm are highlighted by a star. The resulting final confidence intervals are included in the plots.

This exercise was repeated for another data set of size $n = 500$ i.i.d. from a symmetric stable distribution with mean zero and tail index $\alpha = 1.7$. The range of b was there chosen as $b_{small} = 4$ and $b_{big} = 100$.

The plots show that the self-normalizing approach is somewhat less sensitive to the choice of the block size, that is, the confidence interval endpoints change more slowly as b changes. While the plot only shows the results for two data sets, this behavior is typical and was observed for many other simulations as well.

Remark 4.4. Arcones and Giné (1991) considered bootstraping the t-statistic with a smaller bootstrap size m, which corresponds to the block size b of the subsampling method. They suggested to choose $m = n/(\log\log n)^{1+\delta}$ for some small $\delta > 0$. For example, they used $m = 35$ with $n = 50$, and $m = 65$ with $n = 100$. This seems to correspond to $\delta \approx 0.02$.

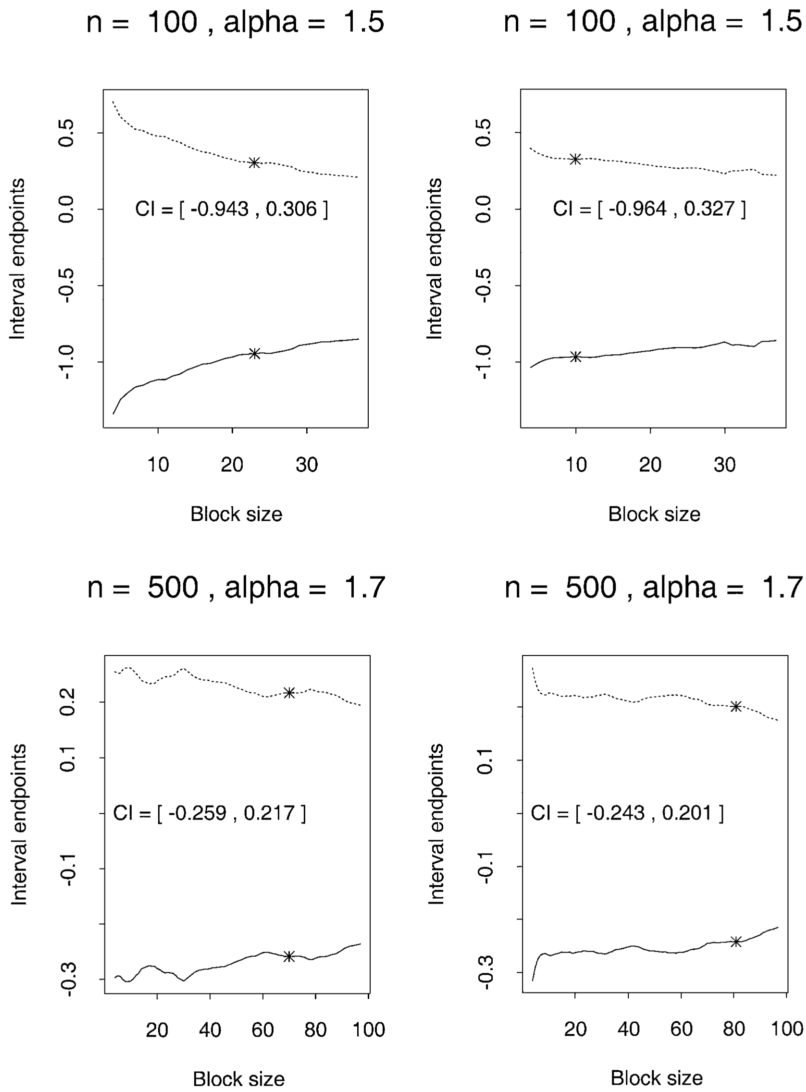

Fig. 1. Illustration of the Minimizing Confidence Interval Volatility Algorithm for two data sets. The plots on the left correspond to the approach appealing to a stable limit, while the plots on the right correspond to the self-normalizing approach. The block sizes selected by the algorithm are highlighted by a star. The final confidence intervals appear within the plots.

5 Small sample performance

The purpose of this Section is to shed some light on the small sample performance of the subsampling method by means of a simulation study. In particular, we want to compare the two approaches of Subsections 3.1 and 3.2. Performance is judged by coverage probabilities of nominal 95% two-sided confidence intervals. We include both equal-tailed and symmetric subsampling intervals (see Section 2) in the study.

The application of the Range estimator $\hat{\alpha}_{I,J}$ requires choices of I, J, the block sizes b_i, and the quantiles t_{2j} and t_{2j-1}; see Bertail et al. (1999). we chose $I = 5$, $J = 10$, the t_{2j-1} equally spaced between 0.01 and 0.25 and $t_{2j} = 1 - t_{2j-1}$, for $j = 1 \dots J$. Finally, the block size b_i was chosen as $n^{0.5\gamma_i}$, rounded to the nearest integer. Here, $\gamma_i = 1 + (\log i/I)/(\log 100)$, for $i = 1 \dots I$. For example, with $I = 5$ and a sample size of $n = 100$, this yields block sizes of 4, 6, 7, 8, and 9.

We consider three data generating mechanisms. First, the stable distribution with mean zero, varying tail index parameter α and varying skewness parameter β. Second, the standard Pareto distribution with distribution function $P\ (X \le x) = 1 - x^{-\alpha}$, for $x > 1$, and varying tail index parameter α; note that the mean of this distribution is given as $\alpha/(\alpha - 1)$ for $\alpha > 0$. Third, a 'symmetrized' Pareto distribution defined by $X = Y - 1$ with probability 0.5 and $X = 1 - Y$ with probability 0.5, where Y has a standard Pareto distribution with tail index parameter α; note this distribution has mean zero. Standard Pareto observations can be easily generated by applying the inverse of the distribution function to Uniform [0,1] observations. For the generation of stable observations, we first used the function *rstab*( ) of the statistical package S-Plus. However, we experienced some problems, since, for skewed distributions ($\beta \neq 0$), *rstab*( ) does not seem to produce variables with the specified mean. In the end, we used the program Stable 2.11 provided by John Nolan at http://www.cas.american.edu/~jpnolan/stable.html.

Our simulation results are based on 1000 repetitions for each scenario. The sample size chosen is always $n = 100$. The model parameters for block size selection Algorithm 4.1 were $b_{small} = 4$, $b_{big} = 30$, $m = 2$, and $k = 2$. Estimated coverage probabilities of nominal 95% confidence intervals are based on 1000 repetitions for each scenario. The results are presented in Table 1. SL stands for the approach of Subsection 3.1, appealing to a Stable Limit, SN stands for the Self-Normalizing approach of Subsection 3.2. The subscripts ET and SYM denote equal-tailed and symmetric intervals, respectively; see Section 2. CLT stands for the Central Limit Theorem approach, falsely assuming a finite variance.

The results for symmetric, stable observations are overall quite satisfactory, although the difference between the equal-tailed and symmetric SN intervals is noteworthy. For skewed, stable observations, coverage decreases with α and this is even more true for Pareto observations. The overall best choice appear to be symmetric SN intervals and while their performance is far from perfect, they present a significant improvement over the CLT intervals. However, it appears that in the context of heavy-tailed observations far bigger sample sizes are needed to achieve overall satisfactory performance as compared to the finite variance case.

6 Summary

In this paper, we have demonstrated that the subsampling method can be used to construct asymptotically correct confidence intervals for the mean when the observations are i.i.d. from a distribution with infinite variance. We proposed two different approaches. The first one is based on the fact that the sample mean, properly standardized, will have a limiting stable law given that the underlying distribution belongs to the normal domain of attraction of a stable

Table 1. Estimated coverage probabilities of various nominal 95% subsampling confidence intervals. The sample size is $n = 100$ always. The estimates are based on 1000 replications for each scenario

Stable observations, $\beta = 0$					
α	SL_{ET}	SL_{SYM}	SN_{ET}	SN_{SYM}	CLT
1.9	0.93	0.92	0.93	0.94	0.94
1.7	0.94	0.95	0.87	0.95	0.94
1.5	0.95	0.94	0.79	0.96	0.92
1.3	0.97	0.96	0.73	0.96	0.98
1.1	0.98	0.98	0.66	0.97	0.98
Stable observations, $\beta = 0.5$					
α	SL_{ET}	SL_{SYM}	SN_{ET}	SN_{SYM}	CLT
1.9	0.93	0.92	0.93	0.94	0.95
1.7	0.94	0.94	0.87	0.95	0.94
1.5	0.94	0.93	0.81	0.94	0.92
1.3	0.89	0.90	0.75	0.89	0.80
1.1	0.52	0.60	0.53	0.59	0.42
Pareto observations					
α	SL_{ET}	SL_{SYM}	SN_{ET}	SN_{SYM}	CLT
1.9	0.82	0.86	0.90	0.92	0.80
1.7	0.82	0.86	0.89	0.91	0.75
1.5	0.74	0.83	0.87	0.88	0.68
1.3	0.61	0.72	0.83	0.82	0.52
1.1	0.41	0.27	0.64	0.61	0.24

law. This approach has the practical disadvantage that the tail index of the underlying distribution has to be estimated. The second approach consists of subsampling the usual t-statistic, which turns out to be a self-normalized sum. It is more general, in the sense that it is not restricted to distributions in the normal domain of attraction of a stable law, and avoids having to estimate the tail index. We proved a theorem that shows the validity of this approach, extending the theory of Politis and Romano (1994).

To deal with the problem of choosing the block size, we proposed an algorithm that minimizes confidence interval volatility over a sensible range of block sizes. Here, volatility is measured by applying a running standard deviation to the confidence interval endpoints in the neighborhood of a particular block size.

We employed a simulation study to examine small sample performance. As to be expected, the results depend on the underlying distribution. Using the second approach, subsampling the t-statistic, to construct symmetric confidence intervals yielded the overall best results.

References

Arcones M (1990) On the asymptotic theory of the bootstrap. Ph.D. dissertation, The City University of New York

Arcones M, Giné E (1989) The bootstrap of the mean with arbitrary bootstrap sample size. Annals of the Institute Henri Poincaré 25:457–481

Arcones M, Giné E (1991) Additions and correction to "the bootstrap of the mean with arbitrary bootstrap sample size". Annals of the Institute Henri Poincaré 27:583–595

Athreya KB (1985) Bootstrap of the mean in the infinite variance case II. Technical Report 86–21, Department of Statistics, Iowa State University

Athreya KB (1987) Bootstrap of the mean in the infinite variance case. Annals of Statistics 15:724–731

Bertail P, Politis DN, Romano JP (1999) On subsampling estimators with unknown rate of convergence. Journal of the American Statistical Association 94:569–579

Bickel P, Götze F, van Zwet WR (1997) Resampling fewer than *n* observations: Gains, losses, and remedies for losses. Statistica Sinica 7:1–31

Chambers JM, Mallows CL, Stuck BW (1976) A method for simulating stable random variables. Journal of the American Statistical Association 71:340–344

deHaan L, Resnick SI (1980) A simple asymptotic estimate for the index of a stable distribution. Journal of the Royal Statistical Society, ser B, 42:83–87

Efron B (1969) Students *t*-test under symmetry conditions. Journal of the American Statistical Association 64:1278–1302

Efron B (1979) Bootstrap methods: Another look at the jackknife. Annals of Statistics 7:1–26

Feller W (1957) An introduction to probability theory and its applications. Vol. 1, 3rd edn, Wiley, New York

Feller W (1971) An introduction to probability theory and its applications. Vol. 2, 2nd edn, Wiley, New York

Hall P (1982) On simple estimates of an exponent of regular variation. Journal of the Royal Statistical Society B 44:37–42

Hall P (1988) On symmetric bootstrap confidence intervals. J. R. Statist. Soc. B, 50:35–45

Hall P (1992) The bootstrap and Edgeworth expansion. Springer-Verlag, New York

Hill B (1975) A simple approach to inference about the tail of the distribution. Annals of Statistics 3:1163–1174

Hotelling H (1961) The behavior of some standard statistical tests under non-standard conditions. Proceedings of the Fourth Berkeley Symposium of Mathematical Statistics Prob. 1:319–360

Kinateder J (1992) An invariance principle applicable to the bootstrap. In: LePage R, Billard L (eds.) Exploring the limits of bootstrap, John Wiley, New York, pp. 157–181

Knight K (1989) On the bootstrap of the sample mean in the infinite variance case. Annals of Statistics 17:1168–1175

Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. Annals of Statistics 17:1217–1241

Lahiri SN (1998) Effects of block lengths on the validity of block resampling methods. Preprint, Department of Statistics, Iowa State University

Logan BF, Mallows CL, Rice SO, Shepp LA (1973) Limit distributions of self-normalized sums. Annals of Probability 1:788–809

Mittnik S, Paolella MS, Rachev ST (1996) A tail estimator for the index of the stable paretian distribution. Working paper Nr. 103, Institute of Statistics and Econometrics, Christian Albrechts University at Kiel

Pickands J (1975) Statistical inference using extreme order statistics. Annals of Statistics 3:119–131

Politis DN, Romano JP (1993) Estimating the distribution of a studentized statistic by subsampling. Bulletin of the International Statistical Institute, 49th Session, Firenze, August 25–September 2, 1993, Book 2, pp. 315–316

Politis DN, Romano JP (1994) Large sample confidence regions based on subsamples under minimal assumptions. Annals of Statistics 22:2031–2050

Politis DN, Romano JP, Wolf M (1997) Subsampling for heteroskedastic time series. Journal of Econometrics 81:281–317

Politis DN, Romano JP, Wolf M (1999) Subsampling. Springer-Verlag, New York

Resnick S (1997) Heavy tail modeling and teletraffic data (with discussion). Annals of Statistics 25:1805–1869

Samorodnitsky G, Taqqu MS (1994) Stable non-Gaussian random processes. Chapman & Hall, New York

Wu J, Carlstein E, Cambanis S (1993) Bootstrapping the sample mean for data with infinite variance. Technical Report No. 296, Department of Statistics, University of North Carolina

Zolotarev VM (1986) One-dimensional stable distributions. Vol. 65 of "Translation of mathematical monographs", American Mathematical Society