Journal of Statistical Planning and Inference 139 (2009) 788 – 801

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

An alternative to the m out of n bootstrap

Hemant Ishwaran[a], Lancelot F. James[b], Mahmoud Zarepour[c,*]

[a]*Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH 44195, USA*
[b]*Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*
[c]*Department of Mathematics and Statistics, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5*

ARTICLE INFO

Article history:
Received 24 November 2006
Received in revised form
24 December 2007
Accepted 28 April 2008
Available online 22 May 2008

MSC:
primary 60G55
60F05
secondary 62E25
62F12

Keywords:
Bayesian bootstrap
Sub-sampling
Stable process
Point process
Dirichlet process

ABSTRACT

It is well known that Efron's bootstrap can fail in settings where the data are heavy tailed and when regularity conditions do not hold. Naturally this applies to weighted bootstrap schemes such as the Bayesian bootstrap. To deal with this, we introduce a Bayesian bootstrap analogue of the m out of n bootstrap. This bootstrap differs from traditional m out of n bootstraps in that all n observations are used in the bootstrap test statistic. Moreover, the method is relatively robust to the selection of m. We establish consistency for the new bootstrap and examine its other useful properties including a connection to the Dirichlet process. Several examples illustrating consistency in settings where the Efron bootstrap fails are given. Further generalizations are suggested.

1. Introduction

Efron's (1979) bootstrap has enjoyed wide popularity since the early 1980s. In the classical setting of n i.i.d. observations, say $X_1,\ldots,X_n$, the idea of the bootstrap is to re-sample these n points with replacement and form a bootstrapped version of the original statistic. Specifically, a bootstrap sample $\{X_1^*,\ldots,X_n^*\}$ is created where the X_i^* are i.i.d. observations from the probability measure

$$P_B(\cdot)=\sum_{i=1}^{n} M_{n,i}\delta_{X_i}(\cdot),$$

where $(M_{n,1},\ldots,M_{n,n})$ follow a multinomial $(n;1/n,\ldots,1/n)$ distribution (δ_x is a probability measure concentrated on x). The bootstrap has been shown to be asymptotically consistent in terms of approximating the sampling distribution of statistics in a variety of settings. This started with the work of Bickel and Freedman (1981). Singh (1981) showed in the case of approximating the sampling mean one could obtain more accurate approximations than those based on the normal distribution. These successes of the bootstrap raised the question whether there are other similar schemes. Rubin (1981) suggested one could mimic many

* Corresponding author.
E-mail address: zarepour@uottawa.ca (M. Zarepour).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.04.032

properties of the bootstrap by replacing the multinomial weights by an n-variate Dirichlet$(1,\ldots,1)$ vector, $(D_1,\ldots,D_n)$. Formally resulting in the Bayesian bootstrap empirical measure,

$$\sum_{i=1}^{n} D_i\delta_{X_i}(\cdot)\overset{\mathrm{d}}{=}\sum_{i=1}^{n}\frac{Z_i}{\sum_{k=1}^{n}Z_k}\delta_{X_i}(\cdot),$$

where $Z_1,\ldots,Z_n$ are i.i.d. exponential (1) random variables.

This procedure is commonly referred to as the Bayesian bootstrap. Lo (1987) subsequently showed the Bayesian bootstrap enjoyed the same consistency properties discussed in Bickel and Freedman. Further, he showed the Bayesian bootstrap was consistent in terms of approximating the posterior distribution of a Dirichlet process, see Ferguson (1973). A generalization of this scheme replaces the n-variate Dirichlet$(1,\ldots,1)$ vector with an n-variate Dirichlet$(\lambda,\ldots,\lambda)$ vector for some $\lambda>0$. This is a sub-class of Bayesian bootstrap clones defined in Lo (1991).

The choice of λ in the Bayesian bootstrap clone is significant. Weng (1989) showed in parallel to Singh (1981), the Bayesian bootstrap (i.e. $\lambda=1$) is second order accurate for approximating the posterior distribution of a Dirichlet mean, but not second order accurate for approximating the distribution of the sample mean. Weng (1989) also showed the choice of $\lambda=4$ leads to a Bayesian bootstrap clone second order accurate for approximating the distribution of the sample mean. See Haeusler et al. (1991), Barbe and Bertail (1995), and James (1997) for the case of more general functionals and weighted bootstrap procedures.

More generally, Mason and Newton (1992) showed replacing the multinomial weights by a variety of exchangeable weights $(W_{n,1},\ldots,W_{n,n})$ satisfying mild conditions would lead to the same consistency results. The work of Giné and Zinn (1984) and Praestgaard and Wellner (1993) on the limiting distribution of bootstrapped and exchangeably weighted bootstrapped empirical measures, as well as numerous investigations by other authors, have shown these schemes are valid in a variety of complex statistical applications. Arguably, among the exchangeable weighted bootstrap procedures, the Bayesian bootstrap is the most appealing. Reasons for this include its ease of use, and its interpretability in relation to the Dirichlet process and empirical likelihood (Owen, 1990). Further, it has interesting usage in terms of phylogenetic trees as described in Efron (2003, p. 139). It may also enjoy some advantages over Efron's bootstrap in situations where ties in the dataset play some role.

However, despite these successes, there are a number of situations where Efron's bootstrap, and hence the exchangeable bootstrap procedures mentioned above, is known to fail asymptotically. There has been a great deal of research interest focused on such examples. Early examples focused on the situation where the data came from a heavy-tailed distribution. For example, see Athreya (1987), Knight (1989), Arcones and Giné (1989), Deheuvels et al. (1993), Zarepour and Knight (1999), Zarepour (1999) and Hall (1990). In these papers, they mostly study asymptotic behavior of the Efron bootstrap applied to the mean and in some cases to the extreme of i.i.d. random variables. Bickel et al. (1997) discuss more general situations where the bootstrap fails unless precise regularity conditions hold. As discussed, and mentioned elsewhere, a widely applicable remedy relies on taking re-samples of the data $\{X_1,\ldots,X_n\}$ of size m rather than n, where m is chosen so that $m=\mathrm{o}(n)$. When m data points are sampled with replacement, resulting in a bootstrap sample $\{X_1^*,\ldots,X_m^*\}$, this scheme is referred to as an m out of n bootstrap. When m is chosen without replacement the scheme corresponds to the sub-sampling procedure of Politis and Romano (1994). See also Wu (1990) and Politis et al. (1999). See the articles by Efron (2003), Davison et al. (2003) and Politis (2003) for a recent overview of some of these concepts.

In this paper, we introduce a Bayesian bootstrap analogue of the m out of n bootstrap. We study the asymptotic behavior of this procedure and demonstrate consistency in examples where Efron's bootstrap fails. Our approach relies on the use of point process theory. We study the limit of the bootstrapped point process and through the use of a continuity argument establish consistency for many types of statistics. Our procedure can be implemented easily in practice and has wide applicability to complex statistics such as convex hulls. It is important to mention the weights introduced in this paper can be an excellent candidate for the generalized bootstrap for estimating equations proposed by Chatterjee and Bose (2005). We also establish an interesting relationship to the Dirichlet process and the important class of finite dimensional Dirichlet processes of Ishwaran and Zarepour (2002). Finally, we emphasize that the work of Del Barrio and Matran (2000) (see also Barbe and Bertail, 1995, pp. 85–91) describes conditions on the weights under which a general weighted bootstrap scheme would be consistent. However, they do not identify an explicit construction such as our Bayesian bootstrap.

2. The Bayesian analogue of the *m* out of *n* bootstrap

We now introduce our Bayesian bootstrap scheme (the Bayesian sub-sampled bootstrap). Let $X=(X_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables with distribution P. Let $(Z_{i,n})_{1\leqslant i\leqslant n}$ be row-wise independent gamma random variables with shape parameter $\lambda m/n$ and scale parameter 1 and independent from X. The Bayesian sub-sampled bootstrap is a special type of Bayesian bootstrap clone (see Lo, 1991 for a definition). Throughout this paper we define

$$S_n=\sum_{i=1}^{n}Z_{i,n}\quad\text{and}\quad W_{i,n}=\frac{Z_{i,n}}{S_n}.$$

The bootstrapped empirical measure for the Bayesian sub-sampled bootstrap is

$$P_W(\cdot) = \sum_{i=1}^{n} W_{i,n}\delta_{X_i}(\cdot) = \sum_{i=1}^{n} \frac{Z_{i,n}}{S_n}\delta_{X_i}(\cdot).$$

The distribution of the vector $(W_{1,n}, \ldots, W_{n,n})$ is Dirichlet with identical parameters $\lambda m/n$. Thus each weight is Beta($\lambda m/n$, $\lambda(m/n)(n-1)$). Perhaps an appropriate name for this procedure is then Bayesian bootstrap m out of n, referred to as Bayesian sub-sampled bootstrap.

One application of the Bayesian subsampled bootstrap is the following procedure for approximating the distribution of a test statistic under the bootstrapped empirical measure:

1. Simulate $Z_{1,n}, \ldots, Z_{n,n}$ independently from a Gamma($\lambda m/n$, 1) distribution.
2. Evaluate $\theta^* = \theta(P_W, \hat{P})$ where $\hat{P}$ is the empirical measure for $X_1, \ldots, X_n$.
3. Repeat 1 and 2, say K times, to obtain $\theta_1^*, \ldots, \theta_K^*$.

One then uses the empirical distribution of $\theta_1^*, \ldots, \theta_K^*$ to approximate the distribution $\mathscr{L}\{\theta(P_W, \hat{P})|X_1, \ldots, X_n\}$ which in turn approximates the distribution $\mathscr{L}\{\theta(\hat{P}, P)\}$.

The Bayesian sub-sampled bootstrap procedure also applies to more complex test statistics. For example, in Section 4 we discuss applications to statistics based on the convex hull of multivariate data. In such instances, the bootstrapped statistic is a complex function involving weights W_i but not expressible in the form of θ^* as aforementioned. Special techniques are needed to handle such cases. Theorem 4.1 of Section 4 discusses the consistency of the Bayesian sub-sampled bootstrap in such settings.

An important feature of the Bayesian sub-sampled bootstrap worth noting is that since the data are not re-sampled, it is not necessary to take m to be an integer as in the m out of n (sub-sampled) bootstrap. Moreover, to have meaningful consistency for the sub-sampled bootstrap we need to have m tending to infinity with n, but as we show in Section 5 this is not necessary for the Bayesian sub-sampled bootstrap. Perhaps this is one of the most notable aspects of the procedure. It is also pointed out in Politis and Romano (1994) that the sub-sampled bootstrap may not work for a statistic sensitive to ties, unless m^2/n tends to 0. Since the Bayesian sub-sampled bootstrap weights are continuous random variables, the problems associated with ties are avoided.

3. Regularly varying tails

First we show our Bayesian bootstrap procedure provides a similar remedy to bootstrap inconsistency as the m out of n bootstrap in regularly varying tail settings. In all cases, m/n tends to 0 as n tends to infinity.

A sequence of i.i.d. random vectors $(X_i)_{i\geqslant 1}$ in $\mathbb{R}^d$ is said to have regularly varying tail if there exists a sequence of positive constants $(a_n)_{n\geqslant 1}$ and a non-null Radon measure μ on $\mathbb{R}^d\backslash\{0\}$ such that

$$nP(a_n^{-1}X_1 \in \cdot) \xrightarrow{\mathrm{v}} \mu(\cdot), \tag{3.1}$$

where the notation $\xrightarrow{\mathrm{v}}$ denotes convergence of measures with respect to the vague topology (Kallenberg, 1983). On $\mathbb{R}$, let $X_1, \ldots, X_n$ be an i.i.d. sample such that

$$H(x) = P(|X_1| > x) = x^{-\alpha}L(x), \tag{3.2}$$

where L is a slowly varying function at infinity (i.e. $L(tx)/L(x)$ tends to 1 as x tends to infinity), $\alpha > 0$ and

$$\lim_{x\to\infty} \frac{P(X_1 > x)}{P(|X_1| > x)} = p, \quad p \in [0,1]. \tag{3.3}$$

We get (3.1) with

$$\mu(\mathrm{d}x) = \alpha(pI(x>0)x^{-\alpha-1} + (1-p)I(x<0)|x|^{-\alpha-1})\,\mathrm{d}x. \tag{3.4}$$

For $\alpha \in (0,2)$, Eqs. (3.2) and (3.3) imply that

$$a_n^{-1}\sum_{i=1}^{n}(X_i - E(X_1 I(|X_1| < a_n))) \to_{\mathrm{d}} \mathscr{S},$$

where $\mathscr{S}$ has a stable distribution and $\to_{\mathrm{d}}$ denotes convergence in distribution. In this case $(X_i)_{i\geqslant 1}$ is said to belong to the domain of attraction of a stable law with index $\alpha \in (0,2)$ (Feller, 1971). In what follows, take $\Gamma_i := E_1 + \cdots + E_i$, where $(E_i)_{i\geqslant 1}$ is an i.i.d. sequence of exponential (1) random variables. Also let $(\Delta_i)_{i\geqslant 1}$ be an i.i.d. sequence of random variables with

$$P(\Delta_1 = 1) = 1 - P(\Delta_1 = -1) = p,$$

which is independent from $(\Gamma_i)_{i\geqslant 1}$. For any given $\alpha>0$, statements (3.2) and (3.3) are equivalent to

$$\sum_{i=1}^{n}\delta_{a_n^{-1}X_i}\rightarrow_{\mathrm{d}}\sum_{i=1}^{\infty}\delta_{\Delta_i\Gamma_i^{-1/\alpha}}, \tag{3.5}$$

where convergence occurs in distribution with respect to the vague topology (Kallenberg, 1983).

In the presence of heavy tailed observations, point process representations play a crucial role in the asymptotic analysis of different statistics. For example, the series representation for a symmetric stable random variable with index $\alpha\in(0,2)$ is the sum of the points of the limiting point process in (3.5), which is

$$\mathscr{S}=\sum_{i=1}^{\infty}\Delta_i\Gamma_i^{-1/\alpha}, \tag{3.6}$$

when $p=q=0.5$. For asymmetric cases when $\alpha\in[1,2)$ a centering term is required for a convergent series but for $\alpha\in(0,1)$ centering is not necessary. See LePage et al. (1981) for more details.

For simplicity, in the following theorem we let $\alpha\in(1,2)$. In cases that $\alpha\in(0,1]$, we can obtain a similar result to Athreya (1987) or Knight (1989) by the use of bootstrap point processes as in Section 4 of this paper. In this paper, without loss of generality, we can assume $\lambda=1$.

Theorem 3.1. *Let $X=(X_i)_{i\geqslant 1}$ be a sequence of i.i.d random variables in the domain of attraction of a non-degenerate α-stable law $\mathscr{S}$, $1<\alpha<2$, i.e.*

$$\mathscr{L}\left\{\sum_{i=1}^{n}(X_i-E(X_1))/a_n\right\}\rightarrow\mathscr{S}.$$

If $\bar{X}=\sum_{i=1}^{n}X_i/n$ then

$$\mathscr{L}\left\{\sum_{i=1}^{n}Z_{i,n}(X_i-\bar{X})/a_{m,n}|X\right\}\overset{\mathrm{P}}{\rightarrow}\mathscr{S},$$

where

$$a_{m,n}=a_m(\Gamma(\alpha))^{1/\alpha}\approx m^{1/\alpha}(\lambda\Gamma(\alpha))^{1/\alpha}.$$

Here $\approx$ denotes asymptotic equivalence as n tends to infinity and $\overset{\mathrm{P}}{\rightarrow}$ denotes convergence in probability.

See the Appendix for the proof.

3.1. Studentization

The above result requires in practice that one knows the index α. A remedy is to use a studentized bootstrap as in Arcones and Giné (1989) expressed in this case as

$$\mathscr{L}\left\{\sum_{i=1}^{n}(Z_{i,n}/S_n)(X_i-\bar{X})\Bigg/\sqrt{\sum_{i=1}^{n}(Z_{i,n}/S_n)^2X_i^2}\,\Bigg|X\right\}
=\mathscr{L}\left\{\sum_{i=1}^{n}Z_{i,n}(X_i-\bar{X})\Bigg/\sqrt{\sum_{i=1}^{n}Z_{i,n}^2X_i^2}\,\Bigg|X\right\}.$$

Consistency follows using Giné (1980) and arguments similar to Arcones and Giné (1989). Unlike the finite variance case, the studentization can be modified in several ways. For example, in the above expression, we can replace the denominator by the maximum of $|Z_{i,n}X_i|$ over $i=1,\ldots,n$. In the symmetric case the limit can be written as

$$T=\sum_{i=1}^{\infty}\Delta_i\Gamma_i^{-1/\alpha}/\Gamma_1^{-1/\alpha},$$

where $p=q=0.5$. See also Zarepour and Knight (1999) and Knight (1989).

3.2. Weighted bootstrap

It is often believed restricting tails of the weights to mimic the tail behavior of the observations may resolve asymptotic inconsistency of weighted bootstraps in heavy tailed cases. Our forthcoming analysis shows such a choice of weights is asymptotically invalid. Consider the point process $\sum_{i=1}^n \delta_{a_n^{-1}X_i}$ and its weighted bootstrap $\sum_{i=1}^n \delta_{a_n^{-1}\mathscr{W}_i X_i}$ where $(\mathscr{W}_i)_{i\geqslant 1}$ is a sequence of i.i.d. random variables. We investigate under what conditions the asymptotic distributions of both point processes are identical. Consider the following theorem:

Theorem 3.2. *Let $X=(X_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables satisfying* (3.2) *and* (3.3). *Let $(a_n)_{n\geqslant 1}$ be the same sequence as defined in* (3.5). *Also let $(\mathscr{W}_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables independent from X. Then*

$$\mathscr{L}\left(\sum_{i=1}^n \delta_{a_n^{-1}\mathscr{W}_i X_i}\middle| X\right)=\mathscr{L}\left(\sum_{i=1}^n \delta_{a_n^{-1}\mathscr{W}_i \Delta_i H^{-1}(\Gamma_i/\Gamma_{n+1})}\middle| X\right) \overset{\text{a.s.}}{\to} \mathscr{L}\left(\sum_{i=1}^\infty \delta_{\mathscr{W}_i\Delta_i\Gamma_i^{-1/\alpha}}\middle| \Delta_1,\Gamma_1,\Delta_2,\Gamma_2,\ldots\right),$$

where H and $(\Delta_i)_{i\geqslant 1}$ are defined in (3.2) *and* (3.5), *respectively*.

Proof. The result immediately follows from Breiman (1968) and an argument similar to Proposition 1 of LePage et al. (1997). See also LePage et al. (1981). □

Theorem 3.2 shows that the limiting distribution of the weighted bootstrap involves $(\mathscr{W}_i)_{i\geqslant 1}$. But for the procedure to be consistent, we must have a limit as on the right-hand side of (3.5), and therefore the weights must disappear. Selecting weights so that they mimic the tail behavior of the data will not resolve this issue. For example, as suggested by Barbe and Bertail (1995), let $(\mathscr{W}_i)_{i\geqslant 1}$ be a sequence of non-negative i.i.d. random variables satisfying (3.1) and independent from X. Also let $\alpha\in(0,2)$ and $G(w)=P(\mathscr{W}_1>w)=w^{-\alpha}L(w)$ for some slowly varying function $L(\cdot)$. Similar to Theorem 3.2, we can show

$$\mathscr{L}\left(\sum_{i=1}^n \delta_{a_n^{-1}\mathscr{W}_i X_i}\middle| X\right)=\mathscr{L}\left(\sum_{i=1}^n \delta_{a_n^{-1}G^{-1}(\Gamma_i/\Gamma_{n+1})X_i}\middle| X\right)\overset{\text{a.s.}}{\to}\mathscr{L}\left(\sum_{i=1}^\infty \delta_{\Gamma_i^{-1/\alpha}X_i}\middle| X\right).$$

Note how the limit depends on X, and therefore that the method is inconsistent.

4. Bayesian bootstrap for point processes

Now we establish asymptotic consistency of the Bayesian bootstrap. Theorem 3.1 and similar results for relatively complex statistics will be established by the use of the continuous mapping theorem. A key ingredient in our analysis is the following lemma:

Lemma 4.1. *Let $(X_{i,n})_{1\leqslant i\leqslant n}$ be a sequence of independent random elements of E (any Polish space) with a probability measure $P_{i,n}$ and μ be a non-null Radon measure on E. Assume that $\sup_{1\leqslant i\leqslant n}P_{i,n}(A)$ tends to zero as n tends to infinity for any compact set A. We have*

$$N_n := \sum_{i=1}^n \delta_{X_{i,n}} \to_{\mathrm{d}} N_\mu$$

if and only if

$$\mu_n := \sum_{i=1}^n P_{i,n} \overset{\mathrm{v}}{\to} \mu, \tag{4.1}$$

where N_μ is a Poisson random measure with a mean measure μ.

Proof. We mimic the proof of Proposition 3.21 of Resnick (1987, p. 154). We show that the Laplace functional of N_n converges to the Laplace functional of N_μ. Let f be a continuous and non-negative function with compact support. Define $f_{i,n}=1-E\,\mathrm{e}^{-f(X_{i,n})}$. We have

$$\log \Psi_{N_n}(f)=\log E\exp(-N_n(f))=\sum_{i=1}^n \log(1-f_{i,n}).$$

Note that $\sup_{1\leqslant i\leqslant n} f_{i,n}$ tends to zero as n tends to infinity as there exists a compact set A (the support of f) such that this supremum is at most $\sup_{1\leqslant i\leqslant n} P_{i,n}(A)$. Using the inequality $\log(1-x)-x\leqslant x^2$ valid on the non-negative halfline, we obtain

$$\left|-\log \Psi_{N_n}(f)-\sum_{i=1}^n f_{i,n}\right| \leqslant \left(\sum_{i=1}^n f_{i,n}\right)^2 \leqslant \sup_{1\leqslant i\leqslant n} f_{i,n}\sum_{i=1}^n f_{i,n}. \tag{4.2}$$

In (4.2), $\sum_{i=1}^n f_{i,n}$ tends to $\int(1-\mathrm{e}^{-f(x)})\,\mathrm{d}\mu(x)$ and the upper bound tends to 0 as n tends to infinity. Therefore the result follows immediately. $\square$

Now we establish asymptotic consistency of our Bayesian bootstrap for point processes induced by observations in the domain of attraction of a stable law with index in $(0,2)$.

Theorem 4.1. *Let $X=(X_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables satisfying* (3.1) *and* (3.2). *Also let $(Z_{i,n})_{1\leqslant i\leqslant n}$ be a sequence of i.i.d. gamma random variables with shape parameter $\lambda m/n$ and $\beta=1$, where $m\to\infty$ and $m=\mathrm{o}(n)$. Also take $a_{m,n}=a_m(\Gamma(\alpha))^{1/\alpha}$. Then given X,*

$$\sum_{i=1}^n \delta_{a_{m,n}^{-1}X_iZ_{i,n}} \to_{\mathrm{d}} \sum_{i=1}^\infty \delta_{\Delta_i\Gamma_i^{-1/\alpha}}$$

in probability where $(\Delta_i)_{i\geqslant 1}$ and $(\Gamma_i)_{i\geqslant 1}$ are defined as before.

Proof. Define $X_{i,n}^*=(X_iZ_{i,n})_{1\leqslant i\leqslant n}$. We prove that $X_{i,n}^*$ satisfies (4.1) with μ as given in (3.4). To prove this we need to show

$$\sum_{i=1}^n P(X_{i,n}^*\in\cdot|X)\overset{\mathrm{v}}{\to}\mu(\cdot)$$

in probability. Take $\lambda=1$ (without loss of generality), $u>0$ and prove

$$\sum_{i=1}^n P(a_{m,n}^{-1}|X_{i,n}^*|>u|X)\overset{\mathrm{P}}{\to}u^{-\alpha} \tag{4.3}$$

and

$$\sum_{i=1}^n P(a_{m,n}^{-1}X_{i,n}^*>u|X)\overset{\mathrm{P}}{\to}pu^{-\alpha}. \tag{4.4}$$

We show that (4.3) holds and (4.4) follows similarly. Following the steps in the proof of Theorem 3.1, we get

$$\begin{aligned}\sum_{i=1}^n E\{P(a_{m,n}^{-1}|X_{i,n}^*|>u|X)\} &= nP(a_{m,n}^{-1}Z_{1,n}|X_1|>u)\\ &= m\int_0^\infty P\left(|X_1|>\frac{u}{x}a_{m,n}\right)\frac{x^{m/n-1}\,\mathrm{e}^{-x}}{\Gamma(m/n+1)}\,\mathrm{d}x.\end{aligned}$$

Since

$$mP(a_{m,n}^{-1}|X_1|>u/x)\to(u/x)^{-\alpha}(\Gamma(\alpha))^{-1},$$

we can write

$$\lim_{n\to\infty} nP(a_{m,n}^{-1}Z_{1,n}|X_1|>u)=u^{-\alpha}\Gamma(\alpha)\lim_{n\to\infty}\int_0^\infty \frac{x^{m/n+\alpha-1}\,\mathrm{e}^{-x}}{\Gamma(m/n+1)}\,\mathrm{d}x.$$

Now it remains to show that

$$n\,\mathrm{Var}\{P(a_{m,n}^{-1}Z_{1,n}|X_1|>u|X)\}\overset{\mathrm{P}}{\to}0.$$

Use Markov's inequality to write

$$P(a_{m,n}^{-1}Z_{1,n}|X_1|>u|X)\leqslant\left(\frac{|X_1|}{ua_{m,n}}\right)^{\alpha/2}E(Z_{1,n}^{\alpha/2}),$$

with

$$\frac{E(Z_{1,n}^{\alpha/2})}{m/n} \to \Gamma(\alpha/2)$$

as n tends to infinity. The result follows from an approach similar to Theorem 3.1. □

4.1. Remarks

1. If we replace $Z_{i,n}$ with weights $W_{i,n}$, where

$$(W_{1,n}, \ldots, W_{n,n}) = \left(\frac{Z_{1,n}}{\sum_{i=1}^{n} Z_{i,n}}, \ldots, \frac{Z_{n,n}}{\sum_{i=1}^{n} Z_{i,n}}\right) \sim \mathrm{Dir}(\lambda m/n, \ldots, \lambda m/n), \tag{4.5}$$

we get the same result as above but we have to rescale the points of the bootstrap process. To see this, notice that

$$\sum_{i=1}^{n} Z_{i,n} \sim \mathrm{Gamma}(\lambda m, 1)$$

and it is easy from (4.5) to see that $\sum_{i=1}^{n} Z_{i,n}/m \xrightarrow{\mathrm{P}} \lambda$.
2. In Theorem 4.1, to achieve almost sure convergence with respect to sample paths, we need to impose both conditions m tends to infinity and $m \log\log n/n$ tends to zero when n tends to infinity. See Zarepour and Knight (1999) for details.
3. We can also show that our weights satisfy conditions $H1, H2, H4$ and $H5^*$ of Del Barrio and Matran (2000). Conditions $H1$–$H2$ are easily established. Now For $H4$, take $\lambda = 1$ without loss of generality and mimic the same technique in Del Barrio and Matran (2000) as follows:

$$P\left(\sqrt{m} \max_{1 \leqslant i \leqslant n} W_{i,n} > \varepsilon\right) \leqslant \frac{n}{m^2 \varepsilon^4} E\{(mW_{1,n})^4\}. \tag{4.6}$$

By plugging in

$$E(W_{1,n}^4) = \frac{m/n(m/n+1)(m/n+2)(m/n+3)}{m(m+1)(m+2)(m+3)}$$

and a simple calculation, one can show easily that the right-hand side of (4.6) converges to zero. Therefore $H4$ holds. It remains to show that

$$m \sum_{i=1}^{n} W_{i,n}^2 \xrightarrow{\mathrm{p}} 1.$$

Since

$$mnE(W_{1,n}^2) = mn \frac{m/n(m/n+1)}{m(m+1)} \to 1,$$

it is enough to show that

$$\mathrm{Var}\left(m \sum_{i=1}^{n} W_{i,n}^2\right) \to 0.$$

Since for all $i \neq j$,

$$m^2 n(n-1) \mathrm{Cov}(W_{i,n}^2, W_{j,n}^2) \to 0,$$

we need to show that

$$m^2 n \,\mathrm{Var}(W_{1,n}^2) = m^2 nE(W_{1,n}^4) - m^2 n(E(W_{1,n}^2))^2 \to 0, \tag{4.7}$$

which easily follows since both terms in (4.7) converge to zero.

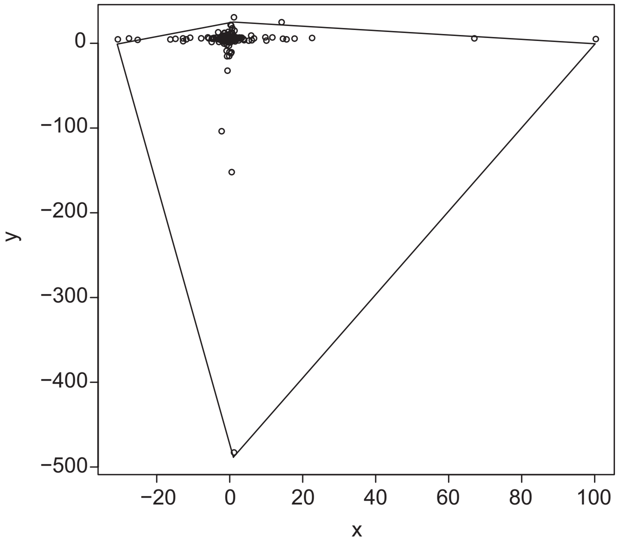

Fig. 1. Convex hull of 200 independent observations with x- and y-coordinates drawn independently from a Cauchy distribution.

4.2. Example 1 (convex hulls)

Theorem 4.1 can be generalized to multivariate distributions with regularly varying tails. Let $X=(X_i)_{i\geqslant 1}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^d$ with the multivariate distribution F and there exists a sequence of positive real numbers $(a_n)_{n\geqslant 1}$ such that

$$nP(a_n^{-1}\|X_1\| \in dr, X_1/\|X_1\| \in d\mathbf{a}) \xrightarrow{v} \mu(dr, d\mathbf{a}) := \alpha r^{-\alpha-1} dr\sigma(d\mathbf{a}), \tag{4.8}$$

where σ is a finite measure defined on the boundary of the d-dimensional unit sphere. For simplicity, in this example, $\|\cdot\|$ denotes the Euclidean norm. Other norms also can be used and the results can be expressed similarly. The convergence in (4.8) holds if F satisfies the multivariate regular variation condition

$$\lim_{t\to\infty} \frac{1-F(tx_1,\ldots,tx_\mathrm{d})}{1-F(t,\ldots,t)} = H(x_1,\ldots,x_d)>0,$$

where

$$H(cx_1,\ldots,cx_d)=c^{-\alpha}H(x_1,\ldots,x_d) \quad \text{for } x_i>0,\ \ i=1,\ldots,d,$$

$c>0$ and $\alpha>0$. For more details, see Resnick (1987, pp. 280–281). In this case it is not difficult to see that

$$\sum_{i=1}^{n} \delta_{a_n^{-1}X_i} \to_{\mathrm{d}} \sum_{i=1}^{\infty} \delta_{U_i\Gamma_i^{-1/\alpha}},$$

where $(\Gamma_i)_{i\geqslant 1}$, as before, are the arrival times of a Poisson process with unit mean and are independent from $(U_i)_{i\geqslant 1}$, an i.i.d. sample taking values on the boundary of the unit sphere in $\mathbb{R}^d$. See Resnick (1987, Section 5.4.2) for more details. Therefore it is easy to see that given X, with the same $(Z_{i,n})_{1\leqslant i\leqslant n}$ (a univariate random variable) and $a_{m,n}$ defined in Theorem 4.1,

$$\sum_{i=1}^{n} \delta_{a_{m,n}^{-1}X_iZ_{i,n}} \to_{\mathrm{d}} \sum_{i=1}^{\infty} \delta_{U_i\Gamma_i^{-1/\alpha}}$$

in probability. The asymptotic analysis of our weighted Bayesian sub-sample bootstrap can be extended to multivariate cases. Some examples are multivariate extremes (coordinatewise extremes) and the convex hulls (the minimal convex set containing the observations) which were not addressed by Del Barrio and Matran (2000). In this case Efron's bootstrap samples of convex hulls remain identical in many cases since few extreme points (usually four points in the independent cases) determine the convex hull (see Fig. 1).

The convex hulls of normalized observations as well as their coordinatewise extremes are continuous functions (with respect to the vague topology) of the point process induced by the normalized observations (Davis et al., 1987, Theorem 3.1). Therefore the continuous mapping theorem implies that the distribution of the convex hull of the normalized bootstrapped sample

$(X_iZ_{i,n}/a_{m,n})_{1\leqslant i\leqslant n}$ converges to that of the convex hull of $(\Gamma_i^{-1/\alpha}U_i)_{i\geqslant 1}$. Convergence holds with respect to the Hausdorff topology (Matheron, 1975) defined on non-empty compact sets. This convergence can be generalized to the number of vertices and continuous functions such as the volume of convex hulls. For example, when $d=2$, this includes the area and perimeter of the convex hull. See Davis et al. (1987) and Zarepour (1999).

4.3. Example 2 (multivariate domain of attraction)

For convenience in notation we consider discussion in $\mathbb{R}^2$. Generalizing our results to any finite dimension is straightforward. Let $(X,Y)=\{(X_i,Y_i)\}_{i\geqslant 1}$ be an i.i.d. sequence of random vectors in $\mathbb{R}^2$. We say (X,Y) is in the domain of attraction of a stable law with indexes $0<\alpha_1,\alpha_2\leqslant 2$ if there exist two positive sequences of constants $(a_n)_{n\geqslant 1}$ and $(b_n)_{n\geqslant 1}$ and two sequences of constants $(c_n)_{n\geqslant 1}$ and $(d_n)_{n\geqslant 1}$ such that

$$\left(\sum_{i=1}^n X_i/a_n - c_n, \sum_{i=1}^n Y_i/b_n - d_n\right) \to_{\mathrm{d}} (\mathscr{S}_1,\mathscr{S}_2)$$

for a random vector $(\mathscr{S}_1,\mathscr{S}_2)$. The marginal limiting distributions are stable with indexes $\alpha_i\in(0,2]$, $i=1,2$. We do not consider the case $\alpha_1=\alpha_2=2$ since the asymptotic validity of the different bootstraps is well known (see also the next example). Similarly when $\alpha_1=2$ and $\alpha_2\in(0,2)$ then $(\mathscr{S}_1,\mathscr{S}_2)$ have to be independent (Resnick and Greenwood, 1979, Theorem 1) and the bootstrap results follow from the univariate case. The only interesting case is when $\alpha_1,\alpha_2\in(0,2)$. In this case let

$$\mathscr{U}_+(x)=\frac{1}{P(X_1>x)},\quad \mathscr{U}_-(x)=\frac{1}{P(X_1\leqslant -x)},\quad x\geqslant 0,\ \ i=1,2.$$

Similarly define $\mathscr{V}_+(y)$ and $\mathscr{V}_-(y)$ by replacing X_1 by Y_1 and x by y. Let

$$\lim_{x\to\infty}\frac{P(X_1>x)}{P(|X_1|>x)}=p,\quad \lim_{y\to\infty}\frac{P(Y_1>y)}{P(|Y_1|>y)}=\pi.$$

Define

$$\mathscr{U}(x)=p\mathscr{U}_+(x)I(x>0)-(1-p)\mathscr{U}_-(-x).$$

Similarly define $\mathscr{V}(y)$ by replacing U by V, x by y, and p by π. We have

$$\sum_{i=1}^n \delta_{(\mathscr{U}(X_i)/n,\mathscr{V}(Y_i)/n)} \to_{\mathrm{d}} \sum_{i=1}^\infty \delta_{\Gamma_i^{-1}U_i},$$

where $\{(U_i,\Gamma_i)\}_{i\geqslant 1}$ are the same as Example 1. For more details see Resnick and Greenwood (1979). Suppose $(Z_{i,n})_{1\leqslant i\leqslant n}$ is the same as in Theorem 3.1. Similar to Example 4.2 we get

$$\sum_{i=1}^n \delta_{Z_{i,n}(\mathscr{U}_+(X_i)/m,\mathscr{V}_-(Y_i)/m)} \to_{\mathrm{d}} \sum_{i=1}^\infty \delta_{\Gamma_i^{-1}U_i}$$

in probability. Now apply the continuous mapping theorem to see that our method remains valid in this case for a point process induced by certain transferred observations. Note that Efron's subsampled bootstrap also works in this case. To see this notice that the bootstrap point process induced by the bootstrap sample $\{(X_i^*,Y_i^*):i=1,2,\ldots,m\}$ is

$$\sum_{i=1}^m \delta_{(a_m^{-1}X_i^*,b_m^{-1}Y_i^*)} \to_{\mathrm{d}} N$$

in probability, which is the same as

$$\sum_{i=1}^m \delta_{(a_m^{-1}X_i,b_m^{-1}Y_i)} \to_{\mathrm{d}} N.$$

The point process N in the above is a Poisson random measure with mean measure ν defined in Theorem 4, part (iii) of Resnick and Greenwood (1979) and $\{(X_i^*,Y_i^*)\}_{(1\leqslant i\leqslant m)}$ are the bootstrap sample of size $m=\mathrm{o}(n)$. This is obvious since

$$mP^*((a_m^{-1}X_i^*,b_m^{-1}Y_i^*)\in\cdot)\xrightarrow{\mathrm{v}}\nu(\cdot).$$

See Zarepour and Knight (1999). The result can be applied to the mapping sum in a similar form as in Theorem 4 of Resnick and Greenwood.

4.4. Example 3 (Shao and Tu, 1995)

Let $(X_i)_{i\geqslant 1}$ be i.i.d. p-dimensional vectors with mean μ and covariance matrix Σ. Let $T_n = g(\bar{X})$ be the desired test statistic where g is a function from $\mathbb{R}^p$ to $\mathbb{R}$ such that g is twice continuously differentiable with $\nabla g(\mu)=0$ and $\nabla^2 g(\mu)\neq 0$. It is well known that Efron's bootstrap fails in this case. Like the subsample bootstrap, the consistency of the Bayesian bootstrap analogue, say $g(\sum_{i=1}^n (Z_{i,n}/S_n)X_i)$, follows from consistency of

$$\sqrt{m}\left(\sum_{i=1}^n (Z_{i,n}/S_n)(X_i-\bar{X})\right)$$

as $m\to\infty$ and $m/n\to 0$. This is immediate from existing results in the literature, say for instance Arenal-Gutiérrez and Matrán (1996). From their work it is also clear that one could define $g(\sum_{i=1}^n W_{i,n}X_i)$, with $(W_{i,n})_{1\leqslant i\leqslant n}$ general exchangeable weights satisfying their conditions $E1$–$E5$, which would be a consistent method when m/n tends to 0 as m,n tend to infinity.

4.5. Example 4 (Bickel and Ren, 1996)

Let $(X_i)_{i\geqslant 1}$ be a sequence of i.i.d. random variables with distribution F. Under double censoring, one only observes the pairs

$$(\mathscr{W}_i,\delta_i)=\begin{cases}(X_i,1), & V_i\leqslant X_i\leqslant Y_i,\\ (Y_i,2), & X_i>Y_i,\\ (V_i,3), & X_i<V_i,\end{cases}$$

where $\{(X_i,Y_i,V_i)\}_{i\geqslant 1}$ are non-negative independent observations with Y_i,V_i corresponding to the right and left censoring variables, respectively. The non-parametric maximum likelihood estimator for $1-F=R$ is defined as a solution to the equation

$$\hat{R}_n(t)=1-Q^{(n)}(t)+\int_{u\leqslant t}\frac{\hat{R}_n(t)}{\hat{R}_n(u)}\,\mathrm{d}Q_2^{(n)}(u)-\int_{u>t}\frac{1-\hat{R}_n(t)}{1-\hat{R}_n(u)}\,\mathrm{d}Q_3^{(n)}(u),$$

where

$$Q_j^{(n)}(t)=\sum_{i=1}^n\frac{1}{n}I\{\mathscr{W}_i\leqslant t,\delta_i=j\}\quad\text{for } j=1,2,3,$$

and

$$Q^{(n)}(t)=\sum_{j=1}^3 Q_j^{(n)}(t).$$

One may desire to test

$$\mathrm{H}_0: F=F_0$$

using the Cramer–von Mises goodness of fit test under double censoring with the Cramer–von Mises statistic defined as

$$\begin{aligned}T_n&=n\int_0^\infty[\hat{F}_n(x)-F_0(x)]^2\,\mathrm{d}F_0(x)\\&=n\int_0^1[\hat{\mathrm{U}}_n-\mathrm{U}]^2\,\mathrm{dU},\end{aligned}$$

where U is the uniform distribution function on [0,1] and

$$\hat{F}_n=1-\hat{R}_n\quad\text{and}\quad\hat{\mathrm{U}}_n=\hat{F}_n\circ F_0^{-1}.$$

Bickel and Ren (1996) propose to use the sub-sample bootstrap to set the critical values for this test. They show Efron's bootstrap fails in this case and that the sub-sample bootstrap procedure is asymptotically consistent and has correct power against $\sqrt{n}$ alternatives. The Bayesian bootstrap may be used in a similar fashion as follows. Define the Bayesian bootstrap analogue of R, say R_n^*, by replacing the $Q_j^{(n)}(t)$ terms with

$$Q_j^*(t)=\sum_{i=1}^n\frac{Z_{i,n}}{S_n}I\{\mathscr{W}_i\leqslant t,\delta_i=j\}\quad\text{for } j=1,2,3.$$

Furthermore, let

$$F_n^* = 1 - R_n^* \quad \text{and} \quad \mathrm{U}_n^* = F_n^* \circ F_0^{-1},$$

and define the Bayesian bootstrap Cramer–von Mises statistic as

$$T_{m,n}^* = m\lambda \int_0^1 [\mathrm{U}_n^*(x) - \mathrm{U}(x)]^2 \, d\mathrm{U}(x).$$

To set critical values for T_n, one uses the Bayesian bootstrap critical values $\gamma_{\alpha,\lambda,m(n)}$, such that

$$P(T_{m,n}^* \geqslant \gamma_{\alpha,\lambda,m(n)}) = \alpha.$$

Since the $Z_{i,n}$'s are continuous random variables, it follows that the choice of $\gamma_{\alpha,\lambda,m(n)}$ is unique. Wellner and Zhan (1996) establish functional central limit theorems for general exchangeable bootstraps of Z-estimators and give as an example the result for double censored data. Using their work combined with some arguments from Bickel and Ren (1996), one can easily show that the Bayesian bootstrap is equal in a first order sense to the sub-sample bootstrap in this case.

5. Some further notes and remarks

Except for unusual cases, the asymptotic failure of Efron's bootstrap can be resolved by $m = o(n)$ resampling. In practice, however, choosing an appropriate m needs careful attention. Asymptotically, $\sqrt{n}$, $\log n$ or $20 \log n$ satisfy the $o(n)$ requirement, but in finite sample settings the actual results can vary dramatically depending on the choice. In what follows, we show our procedures are relatively robust to the selection of m. A valid sub-sample bootstrap requires that m tends to infinity as n tends to infinity. In the next theorem we show that if m does not converge to infinity when n tends to infinity then a reasonable limit is still possible for the Bayesian bootstrap.

In the following theorem we take $\lambda m = v_0$, where m does not depend on n. We denote a gamma measure by $\mathscr{G}(v_0 F)$ and a Dirichlet probability measure by $\mathrm{DP}(v_0 F)$ where v_0 is the concentration measure and F is the prior guess. See Ferguson and Klass (1972) and Ferguson (1973) for more details on Dirichlet and gamma processes and their applications in non-parametric Bayesian inference.

Theorem 5.1. *Let $(X_i)_{i \geqslant 1}$ be a sequence of i.i.d. random variables with a distribution F and let $(T_{i,n})_{1 \leqslant i \leqslant n}$ be a sequence of i.i.d. random variables with a gamma distribution with shape parameter v_0/n and scale parameter 1. Define*

$$\mathrm{N}(x) = v_0 \int_x^\infty \frac{e^{-u}}{u} \, du.$$

(i) *Let $p_{i,n} = T_{i,n} / \sum_{i=1}^n T_{i,n}$ for $i = 1, \ldots, n$. Then*

$$\sum_{i=1}^n \delta_{(T_{i,n}, X_i)} \to_{\mathrm{d}} \sum_{i=1}^\infty \delta_{(N^{-1}(\Gamma_i), X_i)}$$

and

$$\sum_{i=1}^n T_{i,n} \delta_{X_i} \to_{\mathrm{d}} \sum_{i=1}^\infty N^{-1}(\Gamma_i) \delta_{X_i}.$$

In this case, the first limit above is a gamma measure and denoted by $\mathscr{G}(v_0 F)$. Moreover,

$$\sum_{i=1}^n p_{i,n} \delta_{X_i}(\cdot) \to_{\mathrm{d}} \sum_{i=1}^\infty \frac{N^{-1}(\Gamma_i)}{\sum_{i=1}^\infty N^{-1}(\Gamma_i)} \delta_{X_i} \stackrel{\mathrm{d}}{=} \sum_{i=1}^\infty V_i \delta_{X_i} \sim \mathrm{DP}(v_0 F),$$

where

$$V_i = U_i (1 - U_1) \cdots (1 - U_{i-1}) \quad \text{for } i = 1, 2, \ldots \,.$$

Here $(U_i)_{i \geqslant 1}$ is a sequence of i.i.d. random variables with a Beta$(1, v_0)$ *distribution.*

(ii)

$$\mathscr{L}\left(\sum_{i=1}^n \delta_{T_{i,n} X_i} \middle| X \right) \to \mathscr{L}\left(\sum_{i=1}^\infty \delta_{N^{-1}(\Gamma_i) X_i} \middle| X \right)$$

and similarly

$$\mathscr{L}\left(\sum_{i=1}^{n}\delta_{p_{i,n}X_i}\middle|X\right)\to\mathscr{L}\left(\sum_{i=1}^{\infty}\delta_{V_iX_i}\middle|X\right),$$

where $(V_i)_{i\geqslant 1}$ *is the same as in part* (i).

Part (i) of the theorem shows that the bootstrapped empirical measure for the Bayesian bootstrap converges in distribution to a Dirichlet process. In fact, the empirical measure is exactly the finite dimensional prior discussed in Ishwaran and Zarepour (2002). The limiting Dirichlet process of part (i) has concentration parameter v_0 and prior guess F, showing that the limit is a random probability measure concentrated at the true distribution F. The larger the value of v_0, the tighter this concentration. Since v_0 is controlled by m, this shows that a relatively large value of m ensures a reasonable limit; however, it is also clear that the actual choice should not unduly affect inference.

Proof. Using Lemma 4.1 and noting that

$$\sum_{i=1}^{n}P(T_{i,n}\in \mathrm{d}x)=nP(T_{1,n}\in \mathrm{d}x)=\frac{n}{\Gamma(v_0/n)}\mathrm{e}^{-x}x^{v_0/n-1}\,\mathrm{d}x\xrightarrow{\mathrm{v}}v_0\frac{\mathrm{e}^{-x}}{x}\,\mathrm{d}x$$

implies (i). The rest of the theorem follows easily from (i) and the continuous mapping theorem applied to random measures. Also, see Ishwaran and Zarepour (2002) and the references therein for the product representation (stick-breaking representation) of V_i in part (i). □

Remark. Let $(T_{i,n})_{1\leqslant i\leqslant n}$ satisfy the assumptions of Theorem 5.1 and $a_{m,n}$ and $(X_i)_{i\geqslant 1}$ be defined as in Theorem 3.1. The point process $\sum_{i=1}^{n}\delta_{(T_{i,n},a_{m,n}^{-1}X_i)}$ and the random measure $\sum_{i=1}^{n}T_{i,n}\delta_{a_{m,n}^{-1}X_i}$ do not show interesting asymptotic behaviors for our applications. In fact

$$\sum_{i=1}^{n}\delta_{(T_{i,n},a_{m,n}^{-1}X_i)}\to_{\mathrm{d}}\sum_{i=1}^{\infty}\delta_{(N^{-1}(\Gamma_{1,i}),0)}+\sum_{i=1}^{\infty}\delta_{(0,\delta_i\Gamma_{2,i}^{-1/v_0})},$$

where $(\Gamma_{1,i})_{i\geqslant 1}$ and $(\Gamma_{2,i})_{i\geqslant 1}$ are independent and have the same distribution as $(\Gamma_i)_{i\geqslant 1}$. To see this, note from the proof of Theorem 5.1 if $x>0, y>0$,

$$nP(T_{1,n}>x,a_{m,n}^{-1}|X_i|>y)=nP(T_{1,n}>x)P(a_{m,n}^{-1}|X_i|>y)\to N(x)\times 0=0$$

and

$$nP(T_{1,n}>x,a_{m,n}^{-1}|X_i|>0)=nP(T_{1,n}>x)=N(x),$$

where N(x) is defined as in Theorem 5.1.

Acknowledgements

The authors are greatly indebted to the anonymous referee for extremely careful reading of the article, which helped improve the paper significantly. The research of the third author is supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix A. Proof of Theorem 3.1

It follows from Araujo and Giné(1980) (see also Arcones and Giné, 1989) that we only need to check the following conditions:

(A) $\sum_{i=1}^{n}P\{Z_{i,n}|X_i|>\delta a_{m,n}|X\}\xrightarrow{\mathrm{P}}\delta^{-\alpha}$ for $\delta>0$.

(B) $\mathrm{Var}(\sum_{i=1}^{n}(Z_{i,n}X_i/a_{m,n})I\{Z_{i,n}|X_i|\leqslant\delta a_{m,n}\}|X)\xrightarrow{\mathrm{P}}0$.

For (A) we have that

$$\sum_{i=1}^{n}E(P\{Z_{i,n}|X_i|>\delta a_{m,n}|X\})=nP\{Z_{1,n}|X_1|>\delta a_{m,n}\}$$

and

$$nP\{Z_{1,n}|X_1|>\delta a_{m,n}\} = m\int_0^\infty P\left\{|X_1|>\frac{\delta}{u}a_{m,n}\right\}\frac{u^{m/n-1}}{\Gamma(m/n+1)}\,\mathrm{e}^{-u}\,\mathrm{d}u.$$

From (3.1), we get $mP\{|X_1|>ya_{m,n}\}\to y^{-\alpha}[\Gamma(\alpha)]^{-1}$ as m tends to infinity, uniformly in y over relatively compact sets bounded away from 0 (use the metric $d(x,y)=|1/x-1/y|$). Therefore,

$$\begin{aligned}\lim_{m,n\to\infty} nP\{Z_{1,n}|X_1|>\delta a_{m,n}\} &= \delta^{-\alpha}[\Gamma(\alpha)]^{-1}\lim_{m,n\to\infty}\lim_{M\to\infty}\int_{1/M}^{M}\frac{u^{m/n+\alpha-1}}{\Gamma(m/n+1)}\mathrm{e}^{-u}\,\mathrm{d}u\\ &= \delta^{-\alpha}\lim_{m,n\to\infty}\frac{\Gamma(m/n+\alpha)}{\Gamma(m/n+1)\Gamma(\alpha)}\\ &= \delta^{-\alpha}.\end{aligned}$$

The verification of condition (A) is completed by showing

$$n\,\mathrm{Var}(P\{Z_{1,n}|X_1|>\delta a_{m,n}|X\})\xrightarrow{\mathrm{P}}0.$$

Notice that by Markov's inequality

$$P\{Z_{1,n}|X_1|>\delta a_{m,n}|X\}\leqslant\left(\frac{|X_1|}{\delta a_{m,n}}\right)^{\alpha/2}E(Z_{1,n}^{\alpha/2}),$$

with

$$E(Z_1^{\alpha/2})=\frac{m}{n}\frac{\Gamma(m/n+\alpha/2)}{\Gamma(m/n+1)}$$

and

$$\lim_{n\to\infty}\frac{\Gamma(m/n+\alpha/2)}{\Gamma(m/n+1)}=\Gamma(\alpha/2).$$

Thus,

$$nE(P\{Z_{1,n}|X_1|>\delta a_{m,n}|X\})^2\leqslant\delta^{-\alpha}a_{m,n}^{-\alpha}E|X_1|^{\alpha}[E(Z_{1,n}^{\alpha/2})]^2\leqslant\frac{m}{n}K\to 0.$$

For (B), one could proceed using the fact that

$$\lim_{n\to\infty}a_n^{-2}\mathrm{U}(a_n\delta)=\delta^{2-\alpha},$$

where

$$\mathrm{U}(a_n\delta)=E(X_1^2 I\{|X_1|\leqslant\delta a_n\}),$$

and using arguments similar to those above. However, one needs only notice that

$$\begin{aligned}&E\left(\sum_{i=1}^{n}(Z_{i,n}X_i/a_{m,n})^2 I\{Z_{i,n}|X_i|\leqslant\delta a_{m,n}\}|X\right)\\ &\quad\leqslant\delta^{2-\alpha}E\left(\sum_{i=1}^{n}(Z_{i,n}|X_i|/a_{m,n})^{\alpha}I\{Z_{i,n}|X_i|\leqslant\delta a_{m,n}\}|X\right)\\ &\quad\leqslant\delta^{2-\alpha}\left(\sum_{i=1}^{n}|X_i|^{\alpha}/n\right),\end{aligned}$$

which completes the proof. □

References

Araujo, A., Giné, E., 1980. The central limit theorem for real and Banach valued random variables. In: Wiley Series in Probability and Mathematical Statistics.Wiley, New York.

Arcones, M.A., Giné, E., 1989. The bootstrap of the mean with arbitrary bootstrap sample size. Ann. Inst. Henri Poincarè 25, 457–481.

Arenal-Gutiérrez, E., Matrán, C., 1996. A zero-one law approach to the central limit theorem for the weighted bootstrap mean. Ann. Probab. 24, 532–540.

Athreya, K.B., 1987. Bootstrap of the mean in the infinite variance case. Ann. Statist. 15, 724–731.
Barbe, P., Bertail, P., 1995. The Weighted Bootstrap. Springer, New York.
Bickel, P.J., Freedman, D.A., 1981. Some asymptotic theory for the bootstrap. Ann. Statist. 9, 1196–1217.
Bickel, P.J., Ren, J.J., 1996. The *m* out of *n* bootstrap and goodness of fit test with double censored data. In: Reider, H. (Ed.), Robust Statistics Data Analysis and Computer Intensive Methods. Lecture Notes in Statistics. Springer, Berlin.
Bickel, P.J., Goetze, F., van Zwet, W.R., 1997. Resampling fewer than *n* observations: gains, losses and remedies for losses. Statist. Sinica 7, 1–31.
Breiman, L., 1968. Probability. Addison-Wesley, Reading, MA.
Chatterjee, S., Bose, A., 2005. Generalized bootstrap for estimating equations. Ann. Statist. 33 (1), 414–436.
Davis, R.A., Mulrow, E., Resnick, S.I., 1987. The convex hull of a random sample in $\mathbb{R}^d$. Commun. Statist. Stochastic Models 3 (1), 1–27.
Davison, A.C., Hinkley, D.V., Young, G.A., 2003. Recent developments in bootstrap methodology. Silver anniversary of the bootstrap. Statist. Sci. 18, 141–157.
Deheuvels, P., Mason, D.M., Shorack, G.R., 1993. Some results on the influence of extremes on the bootstrap. Ann. Inst. Henri Poincaré 29 (1), 83–103.
Del Barrio, E., Matran, C., 2000. The weighted bootstrap mean for heavy-tailed distributions. J. Theoret. Probab. 13 (2), 1.
Efron, B., 1979. Bootstrap method: another look at the jackknife. Ann. Statist. 7, 1–26.
Efron, B., 2003. Second thoughts on the bootstrap. Silver anniversary of the bootstrap. Statist. Sci. 18, 135–140.
Feller, W., 1971. An Introduction to Probability Theory and Its Applications. vol. II, second ed., Wiley, New York.
Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209–230.
Ferguson, T.S., Klass, M.J., 1972. A representation of independent increment processes without Gaussian component. Ann. Math. Statist. 43, 1634–1643.
Giné, E., 1980. Sums of independent random variables and sums of their squares. In: Proceedings of the seventh Spanish–Portuguese conference on mathematics, Part III (Sant Feliu de Guíxois, 1980). Publ. Sec. Mat. Univ. Autònoma Barcelona 22, pp. 127–132.
Giné, Evarist, Zinn, Joel, 1984. Some limit theorems for empirical processes: With discussion. Ann. Probab. 12 (4), 929–998.
Haeusler, E., Mason, D.M., Newton, M.A., 1991. Weighted bootstrapping of means. CWI Quarterly 4, 213–228.
Hall, P., 1990. Asymptotic properties of the bootstrap for heavy-tailed distributions. Ann. Probab. 18, 1342–1360.
Ishwaran, H., Zarepour, M., 2002. Exact and approximate sum-representations for the Dirichlet process. Canad. J. Statist. 30 (2), 269–283.
James, L.F., 1997. A study of a class of weighted bootstraps for censored data. Ann. Statist. 25, 1595–1621.
Kallenberg, O., 1983. Random Measures. Akademie-Verlag, Berlin.
Knight, K., 1989. On the bootstrap of the sample mean in the infinite variance case. Ann. Statist. 17 (3), 1168–1175.
LePage, R., Woodroofe, M., Zinn, J., 1981. Convergence to a stable distribution via order statistics. Ann. Probab. 9, 624–632.
LePage, R., Podgórski, K., Ryznar, M., 1997. Strong and conditional invariance principles for samples attracted to stable laws. Probab. Theory Related Fields 108, 281–298.
Lo, A.Y., 1987. A large sample study of the Bayesian bootstrap. Ann. Statist. 15, 360–375.
Lo, A.Y., 1991. Bayesian bootstrap clones and a biometry function. Sankhy*ā* Ser. A 53, 320–333.
Matheron, G., 1975. Random Sets and Integral Geometry. Wiley, New York.
Mason, David M., Newton, Michael A., 1992. A rank statistics approach to the consistency of a general bootstrap. Ann. Statist. 20 (3), 1611–1624.
Owen, A., 1990. Empirical likelihood ratio confidence regions. Ann. Statist. 18, 90–120.
Politis, D.N., 2003. The impact of bootstrap methods on time series analysis. Silver anniversary of the bootstrap. Statist. Sci. 18, 219–230.
Politis, D.N., Romano, J.P., 1994. Large sample confidence regions based on subsamples under minimal assumptions. Ann. Statist. 22, 2031–2050.
Praestgaard, J., Wellner, J.A., 1993. Exchangeably weighted bootstraps of the general empirical process. Ann. Probab. 21, 2053–2086.
Politis, Dimitris N., Romano, Joseph P., Wolf, Michael, 1999. Subsampling, Springer Series in Statistics. Springer, New York, 1999.
Resnick, S.I., 1987. Extreme Values, Regular Variation and Point Processes. Springer, New York.
Resnick, S.I., Greenwood, P., 1979. A bivariate stable characterization and domains of attraction. J. Multivariate Anal. 9, 206–221.
Rubin, D., 1981. The Bayesian bootstrap. Ann Statist. 9, 130–134.
Shao, J., Tu, D., 1995. The Jacknife and Bootstrap. In: Springer Series in Statistics.Springer, Berlin.
Singh, K., 1981. On the asymptotic accuracy of Efron's bootstrap. Ann. Statist. 9, 1187–1195.
Wu, C.F.J., 1990. On the asymptotic properties of the jacknife histogram. Ann. Statist. 18, 1438–1452.
Wellner, J.A., Zhan, Y., 1996. Bootstrapping Z–estimators. Technical Report, University of Washington, Seattle.
Weng, C.S., 1989. On a second order asymptotic property of the Bayesian bootstrap mean. Ann. Statist. 17, 705–710.
Zarepour, M., 1999. Bootstrapping convex hulls. Statist. Probab. Lett. 45, 55–63.
Zarepour, M., Knight, K., 1999. Bootstrapping point processes with applications. Stochastic Process. Appl. 84, 81–90.