Generalised bootstrap in non-regular M-estimation problems

Arup Bose[a], Snigdhansu Chatterjee[b,*]

[a] Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 203, B.T. Road, Calcutta 700 035, India
[b] Department of Mathematics, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

Abstract

For estimators of parameters defined as minimisers of $Q(\theta)=Ef(\theta,X)$, we study the asymptotic and generalised bootstrap properties. We concentrate on the case where Q does not have adequate smoothness for standard analysis to work. We describe the properties required by Q as well as bootstrap weights for consistency of the bootstrap.

MSC: primary 62G09; secondary 62G20; 62G35; 62G05; 60F05

Keywords: Nonregular M estimates; Convex function; Bootstrap; Jackknife; Asymptotic distribution

1. Introduction

Let X be a random variable with distribution F and let $f(\theta,X)$ be a measurable function which is convex in θ. Suppose $Q(\theta)=Ef(\theta,X)$ is finite for all θ. Suppose θ_0 is the *unique* minimiser of $Q(\theta)$. This is the unknown parameter to be estimated. By choosing f appropriately, θ_0 can be the mean, any quantile, etc. See Bose (1998) for more examples and details.

Suppose X_i are i.i.d. copies of the random variable X. A natural nonparametric estimate of θ_0 based on these observations is θ_n which minimises the sample quantity

$$Q_n(\theta)=\sum_{i=1}^{n} f(\theta,X_i). \tag{1.1}$$

The above set up can be broadened considerably by allowing f to be a function of the form $f(\theta,X_1,\ldots,X_m)$, defining $Q(\theta)=Ef(\theta,X_1,\ldots,X_m)$, and letting

$$Q_n(\theta)=\sum_{i_1<\cdots<i_m}^{n} f(\theta,X_{i_1},\ldots,X_{i_m}).$$

Examples which fall under this extended set up include variance, U quantiles, L_1 median, Oja median etc. See Bose (1998) for more details.

Thus, this class of M_m estimators is quite rich. Several asymptotic properties of these estimators are known. In particular, these estimators are asymptotically normal under suitable conditions on F and f. See Bose (1998) and the references contained there. A natural question is what would happen when some of these conditions are not valid.

For example, consider the sample median. It has asymptotic normal distribution if F is differentiable with a positive derivative at the population median. Among others, Huang et al. (1996) and Knight (1998a, b) showed that if a suitably weaker condition is used on F, then an appropriately scaled and centred sample median has a nonnormal asymptotic distribution. This condition is an attempt to cover at least some situations where F is not differentiable.

We examine this condition for the median carefully and introduce suitable conditions for the general class of M estimates and call these nonregular M estimates. We show that under these conditions, the nonregular M estimators have possible nonnormal asymptotic distribution. All our results are stated and proved for M_1 estimates. This is done for simplicity. It is possible to formulate similar results for M_m estimates also.

We then turn to the bootstrap approximation results for these estimators. One goal of the bootstrap is to find a consistent, and if possible an accurate, estimate for the distribution of θ_n. It is known from Bose and Chatterjee (2000) that the usual bootstrap and their suitable generalised bootstrap cousins yield distributional consistency for the "regular" M_m estimates. It is also known from the works of Knight (1998a) that for the one dimensional median, the "usual" bootstrap is *not* consistent in nonregular situations while the m out of n bootstrap is consistent. Similar results are available in Knight (1998b) in the regression context.

Following Bose and Chatterjee (2000), we propose that (for $m = 1$), the bootstrap be carried out by minimising

$$Q_{nB}(\theta) = \sum_{i=1}^{n} w_{ni} f(\theta, X_i), \tag{1.2}$$

to obtain the estimator θ_{nB}. This may be termed as generalised bootstrap. If $\mathbf{w}_n = (w_{n1}, \ldots, w_{nn}) \sim$ Multinomial $(n, 1/n, \ldots, 1/n)$, then (1.2) corresponds to the classical bootstrap of Efron (1979). With other choices of distribution of $\mathbf{w}_n$, we obtain the Bayesian bootstrap, the delete-d jackknives for various d, the m out of n bootstrap and several other schemes known in the literature. See Bose and Chatterjee (2000) for a more detailed discussion, specially about the choice of the distribution of $\mathbf{w}_n$. The importance of the study of "generalised bootstrap" stems from trying to understand the salient features of a resampling scheme which are necessary for consistency, higher order accuracy and computational efficiency.

Since the data is i.i.d., a most natural assumption on the weights $(w_{n1}, \ldots, w_{nn})$ is that they form a row exchangeable triangular sequence of nonnegative random variables, independent of the data. Apart from this, some lower order moment assumptions are also made to guarantee consistency. The behaviour of $\sigma_n^2 = \text{Var}(w_{n1})$ turns out to be quite crucial. It is important to note that the delete-d jackknife with d fixed is distributionally inconsistent (see Wu, 1990). For this jackknife, $\sigma_n^2 \to 0$. The classical bootstrap and Bayesian bootstrap are generally consistent in "regular" situations and both of these satisfy $\sigma_n^2 \to 1$. The m out of n bootstrap is known to work for a number of "nonregular problems" and for these $\sigma_n^2 \to \infty$. See Bickel et al. (1997) for examples. Some specific cases are the maximum (Athreya and Fukuchi, 1997a, b; Deheuvels et al., 1993), the sample mean for heavy tailed F (Athreya 1987; Knight, 1989).

Bose and Chatterjee (2000) showed that under suitable conditions on the weights and on the model, the generalised bootstrap is distributionally consistent for M_m estimates. Their assumptions on the model allowed for some nonsmoothness in the function f but along with other conditions it was assumed that $Q(\theta)$ is twice differentiable and the matrix of second derivatives at θ_0 is positive definite.

When specialised to the sample median, this condition is the same as the standard assumption that F is differentiable with a positive density at the median. Thus, the case of nondifferentiable F was excluded.

Knight (1998a) demonstrates that under his alternate nonregular condition for the median, the usual bootstrap is inconsistent but the m out of n bootstrap is consistent.

We investigate the consistency of the generalised bootstrap for nonregular M estimates. Our study on bootstrap in the nonregular cases identifies which schemes are consistent and which are not. The schemes with $\sigma_n^2 \to \infty$ turn out to be consistent and the others turn out to be inconsistent. In particular, the usual bootstrap and the delete d jackknives for fixed d are inconsistent. The m out of n bootstrap is consistent only if $m/n \to 0$.

In Section 2, we give result on the asymptotic distribution of the M_1 estimates under nonregularity and its proof. In Section 3, we study the asymptotics of the bootstrap. In Section 4, we give some examples and report results of a small simulation exercise. The simulation exercise shows how different resampling techniques perform in nonregular cases. The simulation illustrates that there are various possible choices of consistent resampling techniques for which $\sigma_n^2 \to \infty$, and thus the choice of a proper technique is an important issue and merits future investigation.

2. Asymptotic distribution of M_1 estimates

The assumption which defines the nonregular behaviour is the following.

Assumption A. (i) The minimiser θ_0 of $Q(\theta)$ is unique. Further, there exists a $\Psi(u)$, and a nondecreasing sequence $\{a_n\}$ such that the following limit exists:

$$\lim_{n\to\infty} n^{1/2} a_n [Q(\theta_0 + a_n^{-1}u) - Q(\theta_0)] = \Psi(u).$$

(ii) For almost every x (with respect to the Lebesgue measure), $u'x + \Psi(u)$ has a unique minimum.

Remark 1. Note that since f is convex, Q is also convex. It follows that $\Psi(\cdot)$ is also convex. It is well known that if a sequence of convex functions converge pointwise, then the convergence is uniform on compact sets. See for example Rockafellar (1970). Hence the convergence in Assumption A (i) is uniform over compact subsets of Θ.

Remark 2. Every convex function C has a measurable subgradient which we denote by ∇C. This subgradient satisfies

$$C(u_1) + (u_2 - u_1)^{\mathrm{T}} \nabla C(u_1) \leqslant C(u_2) \tag{2.1}$$

for all u_1, u_2 in the domain of C. Further, this subgradient is nondecreasing in the sense that $(u_1 - u_2)'(\nabla C(u_1) - \nabla C(u_2)) \geqslant 0$ for all u_1 and u_2. If $g(u,x)$ is the subgradient of $f(u,x)$ (g can be chosen to be measurable in both arguments) then $E(g(\theta,X)) = \nabla Q(\theta)$ (see Lemma 2.3 later). It is easy to see that since θ_0 is the unique minimiser, $\nabla Q(\theta_0) = 0$ (see Lemma 2.3 later). Further, if Q is twice differentiable, then typically $a_n = n^{-1/2}$ and $\Psi(u) = u'\nabla Q(\theta_0)u/2$. This corresponds to the regular case.

Knight (1998a, b, 1999) has studied the median under nonregular conditions. In that case we can take $f(\theta, x) = |x - \theta|$, see Section 4.

We now state the asymptotic distribution theorem for the nonregular M estimates. As pointed out earlier, an appropriate version of this theorem also holds for the M_m estimates.

Theorem 2.1. *Under Assumption A* (i) *and* (ii) *above*, $a_n(\theta_n - \theta_0) \overset{\mathscr{D}}{\Rightarrow} \arg\min Z(u)$, *where* $Z(u) = u^{\mathrm{T}} W + \Psi(u)$ *and* $W \sim \mathrm{N}(0, \Sigma)$ *with* $\Sigma = Eg(\theta_0, X)g(\theta_0, X)^{\mathrm{T}}$.

Note that $\nabla\Psi$, being the gradient of a convex function, is nondecreasing. If it is one–one, then it is strictly increasing and its inverse is well defined. In this case, $a_n(\theta_n - \theta_0) \stackrel{\mathscr{D}}{\Rightarrow} (\nabla\Psi)^{-1}(-W)$ where $W \sim \mathrm{N}(0,\Sigma)$ with $\Sigma = Eg(\theta_0, X)g(\theta_0, X)^{\mathrm{T}}$. Knight (1998a) has an example where $\nabla\Psi$ is not strictly increasing, but nevertheless, the limit distribution of the median exists.

Before we give the proof of this theorem, let us quote two results that we use in the proof and also in the proof of the bootstrap theorem given later. The first is a stochastic version of the uniform convergence result remarked above. This has been used by Haberman (1989), Niemiro (1992) and Pollard (1988). We quote this result.

Lemma 2.1. *Suppose $\{A_n(s)\}$ is a sequence of convex random functions defined on an open set $S \in \mathbb{R}^p$, which converges in probability to some $A(s)$, for each S. Then $\sup_{s\in K} |A_n(s) - A(s)| \stackrel{\mathrm{P}}{\to} 0$ for each compact subset K of S.*

The following result follows from the works of Hjort and Pollard (1993) and Knight (1998a).

Lemma 2.2. *Suppose that $Z_n(u)$ is a sequence of random convex functions with minimisers u_n. Suppose that the finite dimensional distributions of $Z_n(u)$ converge in distribution to those of a random function $Z(u)$ which has a unique minimum u. Then $u_n \stackrel{\mathscr{D}}{\Rightarrow} u$.*

The following lemma is useful in the sequel. Let $g(u,x)$ be a subgradient of $f(u,x)$.

Lemma 2.3. (a) *For every θ, $\nabla Q(\theta) = Eg(\theta, X)$. Further, $Eg(\theta_0, X) = 0$.*

(b) *For any fixed u, $u^{\mathrm{T}}Eg(\theta_0 + u/a_n, X) \downarrow 0$. Consequently, $u^{\mathrm{T}}E[g(\theta_0 + u/a_n, X) - g(\theta_0, X)] \downarrow 0$, and hence $u'Y_n(u) = u^{\mathrm{T}}[g(\theta_0 + u/a_n, X) - g(\theta_0, X)] \to 0$ almost surely.*

Proof. First note that from Assumption A(i), for every fixed u,

$$\lim_{n\to\infty} a_n[Q(\theta_0 + a_n^{-1}u) - Q(\theta_0)] = 0. \tag{2.2}$$

Using (2.1),

$$-(f(\theta_0 - u/a_n, X) - f(\theta_0, X)) \leqslant a_n^{-1}u^{\mathrm{T}}g(\theta_0, X) \leqslant (f(\theta_0 + u/a_n, X) - f(\theta_0, X)).$$

Now take expectations, to get

$$-a_n(Q(\theta_0 - u/a_n) - Q(\theta_0)) \leqslant u^{\mathrm{T}}Eg(\theta_0, X) \leqslant a_n(Q(\theta_0 + u/a_n) - Q(\theta_0)).$$

Now using (2.2) the second part of (a) follows. To establish the first part from the differential and argue as above. We omit the details.

For (b) we follow similar steps.

We again use (2.1) to obtain

$$\begin{aligned}-(f(\theta_0 - u/a_n, X) - f(\theta_0, X)) &\leqslant a_n^{-1}u^{\mathrm{T}}g(\theta_0 + u/a_n, X)\\ &\leqslant (f(\theta_0 + 2u/a_n, X) - f(\theta_0 + u/a_n, X))\end{aligned}$$

and again by taking expectations, we have

$$\begin{aligned}-a_n(Q(\theta_0 - u/a_n) - Q(\theta_0)) &\leqslant u^{\mathrm{T}}Eg(\theta_0 + u/a_n, X)\\ &\leqslant a_n(Q(\theta_0 + 2u/a_n) - Q(\theta_0)) - a_n(Q(\theta_0 + u/a_n) - Q(\theta_0))\end{aligned}$$

and we get the first part of (b) from (2.2).

Again by using (2.1), and the assumption that a_n is nondecreasing, it follows that for every fixed u the sequence $u^{\mathrm{T}}Y_n(u)$ is nonnegative and decreasing. Denote the nonnegative limit random variable by $Y(u)$. From (a) and the first part of (b) it follows that $EY(u)=0$, consequently $Y(u)=0$ almost surely. This proves the second part of (b). □

Proof of Theorem 2.1. Define

$$Z_n(u)=a_n n^{-1/2}\sum_{i=1}^{n}[f(\theta_0+a_n^{-1}u,X_i)-f(\theta_0,X_i)].$$

Since f is convex, $Z_n(u)$ is also convex. Also note that if θ_n is a minimiser of $\sum f(\theta,X_i)$, then $u_n=a_n(\theta_n-\theta_0)$ is a minimiser of $Z_n(\cdot)$. We point out here that our minimisers, here and elsewhere are chosen in a measurable way. This is possible by a suitable selection theorem. See, for example, Castaing and Valadier (1977) or Niemiro (1992).

We can write $Z_n(u)=T_{n1}(u)+T_{n2}(u)+T_{n3}(u)$ where

$$T_{n1}(u)=a_n n^{-1/2}\sum_{i=1}^{n}[f(\theta_0+a_n^{-1}u,X_i)-f(\theta_0,X_i)-a_n^{-1}u^{\mathrm{T}}g(\theta_0,X_i)]$$

$$-n^{1/2}a_n[Q(\theta_0+a_n^{-1}u)-Q(\theta_0)]$$

$$T_{n2}(u)=n^{1/2}a_n[Q(\theta_0+a_n^{-1}u)-Q(\theta_0)]$$

$$T_{n3}(u)=n^{-1/2}\sum_{i=1}^{n}u^{\mathrm{T}}g(\theta_0,X_i).$$

We now study these three terms.

(i) First note that from the assumptions we have that $T_{n2}(u)\to\Psi(u)$, in fact uniformly over compact sets by convexity.

(ii) Let us write $T_{n1}(u)=\sum_{i=1}^{n}Z_{ni}(u)$, where

$$Z_{ni}(u)=a_n n^{-1/2}[f(\theta_0+a_n^{-1}u,X_i)-f(\theta_0,X_i)-a_n^{-1}u^{\mathrm{T}}g(\theta_0,X_i)]$$

$$-n^{-1/2}a_n[Q(\theta_0+a_n^{-1}u)-Q(\theta_0)].$$

Then we have $EZ_{ni}(u)=0$ for all u and by using independence, $\mathrm{Var}(T_{n1}(u))=\mathrm{Var}(\sum_{i=1}^{n}Z_{ni}(u))=nEZ_{n1}^2(u)$. Now

$$nEZ_{n1}^2(u)=a_n^2\mathrm{Var}[f(\theta_0+a_n^{-1}u,X_1)-f(\theta_0,X_1)-a_n^{-1}u^{\mathrm{T}}g(\theta_0,X_1)]$$

$$\leqslant a_n^2E[f(\theta_0+a_n^{-1}u,X_1)-f(\theta_0,X_1)-a_n^{-1}u^{\mathrm{T}}g(\theta_0,X_1)]^2.$$

On the other hand, since f is convex, we have

$$0\leqslant f(\theta_0+a_n^{-1}u,X_1)-f(\theta_0,X_1)-a_n^{-1}u^{\mathrm{T}}g(\theta_0,X_1)$$

$$\leqslant a_n^{-1}u^{\mathrm{T}}(g(\theta_0+a_n^{-1}u,X_1)-g(\theta_0,X_1))=a_n^{-1}u^{\mathrm{T}}Y_n(u)\quad\text{say}.$$

Thus $\mathrm{Var}(T_{n1}(u))\leqslant E[u^{\mathrm{T}}(g(\theta_0+a_n^{-1}u,X_1)-g(\theta_0,X_1))]^2=E(u^{\mathrm{T}}Y_n(u))^2$.

Now using Lemma 2.3(b), second part, it follows that $T_{n1}(u)\xrightarrow{\mathrm{P}}0$ for each fixed u. Now note that by (i) above, the second term of T_{n1} converges. Hence the first term also converges. Now by convexity of each term in T_{n1}, this convergence is uniform on compact sets by Lemma 2.1. Thus we can conclude that $T_{n1}(u)\xrightarrow{\mathrm{P}}0$ uniformly on compact sets of u.

(iii) Since θ_0 is the unique minimiser of $Q(\theta)$, it easily follows that $Eg(\theta_0, X) = 0$. Hence by central limit theorem, $T_{n3}(u) = u^{\mathrm{T}} W_n \overset{\mathscr{D}}{\Rightarrow} u^{\mathrm{T}} W$, where $W \sim \mathrm{N}(0, \Sigma)$ with $\Sigma = Eg(\theta_0, X) g(\theta_0, X)^{\mathrm{T}}$.

Define $Z(u) = u^{\mathrm{T}} W + \Psi(u)$. From the above steps (i)–(iii), it follows that the finite dimensional distributions of $Z_n(u)$ converge to those of $Z(u)$. Now we can apply Lemma 2.2(i) to claim that

$$u_n = a_n(\theta_n - \theta_0) \overset{\mathscr{D}}{\Rightarrow} \arg\min Z(\cdot). \qquad \square$$

3. Bootstrap asymptotics of M_1 estimates

We first discuss the various conditions we impose on the bootstrap weights. Let $\{w_{ni},\ i = 1, \ldots, n,\ n = 1, 2, \ldots\}$ be a triangular sequence of nonnegative, row-wise exchangeable random variables, independent of $\{X_1, \ldots, X_n \ldots\}$. We use the notations P_{B}, E_{B}, V_{B} to, respectively, denote probabilities, expectations and variance with respect to the bootstrap probabilities, that is, with respect to the distribution of the weights, conditional on the given data $\{X_1, \ldots, X_n, \ldots\}$. Let $\sigma_n^2 = V_{\mathrm{B}} w_{n1}$, $W_{ni} = \sigma_n^{-1}(w_{ni} - 1)$. The notations k and K will be used to denote generic constants. The following conditions (3.1)–(3.3) on the weights are assumed:

$$E_{\mathrm{B}} w_{n1} = 1, \tag{3.1}$$

$$0 < k_1 < \sigma_n^2, \tag{3.2}$$

$$c_{11} = \mathrm{corr}(w_{ni}, w_{nj}) = \mathrm{O}(n^{-1}). \tag{3.3}$$

It may be noted that whenever $\sum_i^n w_{ni} = C_n$ for some constant K_n, then condition (3.3) is automatically satisfied.

Not that for the m out of n bootstrap, $\sigma_n^2 = n/m$. It is known from earlier works that for this bootstrap to be consistent (in probability or almost surely) in nonregular situations, some conditions are needed on the resample size m. For our results we make the following assumption.

Assumption B. There exists a function Ψ_1 (which is necessarily convex) for which $\nabla \Psi_1$ is one and a sequence β_n such that

$$n^{1/2} \beta_n \sigma_n^{-1} [Q(\theta_0 + \beta_n^{-1} u) - Q(\theta_0)] \to \Psi_1(u).$$

Note that if $\beta_n = a_n$ and $\sigma_n \to 1$ then this is exactly Assumption A (with $\Psi = \Psi_1$). This will include the usual bootstrap. Further, if for some subsequence m of n, $\beta_n = a_m$ and $m^{-1/2} n^{1/2} \sigma_n^{-1} \to 1$ then Assumption B holds if Assumption A holds (with $\Psi = \Psi_1$). This will include the m out of n bootstrap.

Let $G_{n\mathrm{B}0}$ and $G_{n\mathrm{B}}$ be the probability distributions of $\theta_{n\mathrm{B}}$ with the following centering and normings:

$$G_{n\mathrm{B}0}(x) = P_{\mathrm{B}}[\beta_n(\theta_{n\mathrm{B}} - \theta_0) \leqslant x].$$

$$G_{n\mathrm{B}}(x) = P_{\mathrm{B}}[\beta_n(\theta_{n\mathrm{B}} - \theta_n) \leqslant x].$$

Let W and W_{B} be two independent copies of a random variable having distribution $\mathrm{N}(0, \Sigma)$ where $\Sigma = Eg(\theta_0, X) g(\theta_0, X)^{\mathrm{T}}$. Let C and C_1 be positive constants where C can be ∞. Let $G_{\mathrm{B}W0}(x)$ and $G_{\mathrm{B}W}(x)$ be the random probabilities (given W)

$$G_{\mathrm{B}W0} = P[(\nabla \Psi_1)^{-1}(-W_{\mathrm{B}} - C^{-1} W) \leqslant x],$$

$$G_{\mathrm{B}W} = P[(\nabla \Psi_1)^{-1}(-W_{\mathrm{B}} - C^{-1} W) - C_1 (\nabla \Psi)^{-1}(-W) \leqslant x].$$

Let $U = (\nabla \Psi^{-1})(-W)$.

We now state the bootstrap theorem.

Theorem 3.1. *Suppose Assumption A(i)–(iii) hold. Further, the weight sequence is such that (3.1)–(3.3) hold, Assumption B holds and a bootstrap central limit theorem is satisfied. Assume that $\sigma_n^2 \to C^2$. Then, for every $x_1,\ldots,x_k$,*

$$(a_n(\theta_n - \theta_0), G_{nB0}(x_1), \ldots, G_{nB0}(x_k)) \overset{\mathscr{D}}{\Rightarrow} (U, G_{BW0}(x_1), \ldots, G_{BW0}(x_k)).$$

If further $\beta_n/a_n \to C_1$, then

$$(a_n(\theta_n - \theta_0), G_{nB}(x_1), \ldots, G_{nB}(x_k)) \overset{\mathscr{D}}{\Rightarrow} (U, G_{BW}(x_1), \ldots, G_{BW}(x_k)).$$

Let us examine the consequences of the above result. The generalised bootstrap is distributionally consistent if and only if the random distribution $G_{BW}(x)$ is the same as the nonrandom distribution function $G(x) = P[(\nabla\Psi)^{-1}(-W) \leqslant x]$. Note that this is not possible if $(\nabla\Psi)^{-1}$ and $(\nabla\Psi_1)^{-1}$ are unequal.

If $(\nabla\Psi)^{-1}(u) = (\nabla\Psi_1)^{-1}(u) = \lambda u$, then $G(x) = G_{BW}(x)$ and the bootstrap is consistent (provided we choose β_n to be a_n/σ_n or asymptotically equivalent to it).

If $(\nabla\Psi)(u) = (\nabla\Psi_1)^{-1}(u)$ but they are NOT linear, then it is easy to see that if C or C_1^{-1} is finite, the bootstrap is inconsistent. This includes the usual bootstrap, m out of n bootstrap with m/n tending to a finite limit and the delete-d jackknife with $d \to \infty$ such that d/n tends to a non zero limit.

However, if $(\nabla\Psi)^{-1}(u) = (\nabla\Psi_1)^{-1}(u)$, and both $C = \infty$ and $C_1 = 0$, then the bootstrap becomes consistent. This includes the m out of n bootstrap with $n/m \to \infty$.

The question of accuracy of a consistent bootstrap approximation is very relevant. However, we do not address this issue here. It may be conjectured that the speed of convergence of the bootstrap distribution will depend, among other things, on the behaviour of σ_n^2 as $n \to \infty$. A more detailed study on the accuracy aspect can lead to criteria for choice of m in the m out of n bootstrap.

The following result of Knight (1998a) is used for proving Theorem 3.1.

Lemma 3.1. *Suppose that $\tilde{Z}_n(u)$ and $\{Z_n\}$ are random convex functions such that*

(a) *for any compact set $\mathbb{K}$, $\sup_{u\in\mathbb{K}} |\tilde{Z}_n(u) - Z_n(u)| \xrightarrow{\mathrm{P}} 0$;*

(b) *$\tilde{Z}_n(u)$ has an almost sure unique minimiser V_n which is $\mathrm{O_P}(1)$;*

(c) *for each $\delta > 0$, $H_n(\delta) = [\inf_{|u-V_n|=\delta} |Z_n(u) - \tilde{Z}_n(V_n)|]^{-1} = \mathrm{O_P}(1)$,*

Then if U_n minimises $Z_n U_n - V_n \xrightarrow{\mathrm{P}} 0$.

Remark 3. It may be pointed out that condition (c) above holds if the finite dimensional distributions of $Z_n(u)$ converge to that of $Z(u)$ for some $Z(\cdot)$, see Knight (1998a).

Proof of Theorem 3.1. Note that we have assumed $\sigma_n^2 \to C^2$, where $C = \infty$ is a possibility. Define

$$Z_{nB}(u) = \beta_n n^{-1/2} \sigma_n^{-1} \sum_{i=1}^{n} w_{ni}[f(\theta_0 + \beta_n^{-1}u, X_i) - f(\theta_0, X_i)].$$

Since w_{ni} are nonnegative, $Z_{nB}(u)$ is convex. Also note that if θ_{nB} is a minimiser of $\sum w_{ni} f(\theta, X_i)$, then $u_{nB} = \beta_n^{-1}(\theta_{nB} - \theta_0)$ is a minimiser of $Z_{nB}(\cdot)$.

We can write $Z_{nB}(u) = T_{nB1} + T_{nB2} + T_{nB3} + T_{nB4}$ where

$$T_{nB1} = \beta_n n^{-1/2} \sigma_n^{-1} \sum_{i=1}^{n} w_{ni}[f(\theta_0 + \beta_n^{-1}u, X_i) - f(\theta_0, X_i) - \beta_n^{-1} u^{\mathrm{T}} g(\theta_0, X_i)]$$

$$- n^{-1/2} \beta_n \sigma_n^{-1} [Q(\theta_0 + \beta_n^{-1}u) - Q(\theta_0)] \sum_{i=1}^{n} w_{ni},$$

$$T_{nB2} = n^{1/2}\beta_n\sigma_n^{-1}[Q(\theta_0 + a_n^{-1}u) - Q(\theta_0)]\left(n^{-1}\sum_{i=1}^{n} w_{ni}\right),$$

$$T_{nB3} = n^{-1/2}\sum_{i=1}^{n} W_{ni}u^{\mathrm{T}}g(\theta_0, X_i) = u'W_{nB} \quad \text{say},$$

$$T_{nB4} = n^{-1/2}\sigma_n^{-1}\sum_{i=1}^{n} u^{\mathrm{T}}g(\theta_0, X_i) = u'\sigma_n^{-1}W_n \quad \text{say}.$$

(i) First note that the assumptions on the weights implies that $(n^{-1}\sum_{i=1}^{n} w_{ni}) \to 1$ in P_{B} probability. Using this fact and the same techniques as in the proof of Theorem 2.1, we can show that for any compact set K, and any $\varepsilon > 0$, under assumptions $P_{\mathrm{B}}[\sup_{u\in K}|T_{nB1}(u)| > \varepsilon] \xrightarrow{\mathrm{p}} 0$. We omit the details.

(ii) By the observation on the weights in (i) above and Assumption B, it follows that $T_{nB2} \to \Psi_1(u)$ uniformly over compact sets, in probability.

(iii) By using central limit theorem for weighted sums of exchangeable random variable sequences (see Praestgaard and Wellner, 1993; Arenal-Gutierrez and Matran, 1996 or Bose and Chatterjee, 2000), it is easy to see that conditional on the data, $T_{nB3} \xrightarrow{\mathscr{D}} u^{\mathrm{T}}W_{\mathrm{B}}$, almost surely (P), where $W_{\mathrm{B}} \sim \mathrm{N}(0, \Sigma)$ with $\Sigma = Eg(\theta_0, X)$ $g(\theta_0, X)^{\mathrm{T}}$. Also note that unconditionally, by CLT, $W_n \xrightarrow{\mathscr{D}} W$ where W has the same distribution as W_{B} above. Moreover W and W_{B} are independent.

We now use Lemma 3.1. Define $\tilde{Z}_{nB}(u) = u^{\mathrm{T}}(W_{nB} + \sigma_n^{-1}W_n) + \Psi_1(u)$. We shall show that the conditions of the lemma are satisfied in the conditional world (in probability).

(a) From the arguments given in (i) and (ii), it follows that for every $\varepsilon > 0$, $P_{\mathrm{B}}[\sup_{u\in K}|Z_{nB}(u) - \tilde{Z}_{nB}(u)| > \varepsilon] \xrightarrow{\mathrm{p}} 0$, verifying condition (a) of Lemma 3.1.

(b) By assumption (ii), $\tilde{Z}_{nB}(u)$ has a unique minimiser, say V_n which is obtained by solving, $V_n = (\nabla\Psi_1)^{-1}$ $(-W_{nB} - \sigma_n^{-1}W_n)$. Since W_{nB} converges in distribution conditionally and W_n converges in distribution unconditionally, it follows that given $\varepsilon > 0$, there exists $K > 0$ such that $P[P_{\mathrm{B}}[|V_n| > K] > \varepsilon] < \varepsilon$. This implies that $V_n = \mathrm{O}_{P\mathrm{B}}(1)$ with as large probability as we please, verifying condition (b) of Lemma 3.1 (with large probability).

(c) We now verify condition (c) of Lemma 3.1. By the Remark after the statement of the lemma, it is enough to show that the finite dimensional distributions of Z_{nB} converge. But since Z_{nB} is approximated by $\tilde{Z}_{nB}$ uniformly over compact sets, it is enough to verify this for the latter process. Fix $u_1, \ldots, u_k$ and consider $P_{\mathrm{B}}[\tilde{Z}_{nB}(u_1) \leqslant x_1, \ldots, \tilde{Z}_{nB}(u_k) \leqslant x_k)]$. It is easy to see, by using (iii) above that the above probability converges in distribution to the conditional probability (hold W fixed) $P[u_1^{\mathrm{T}}(W_{\mathrm{B}} + C^{-1}W) + \Psi_1(u_1) \leqslant x_1, \ldots, u_k^{\mathrm{T}}(W_{\mathrm{B}} + C^{-1}W) + \Psi_1(u_k) \leqslant x_k]$. This verifies (c) (in distribution).

Now the first part of the Theorem follows from the above discussion. The second part of the Theorem follows by writing $\beta(\theta_{nB} - \theta_n) = \beta(\theta_{nB} - \theta_0) - (\beta_n/a_n)a_n(\theta_n - \theta_0)$ and using the assumption that $\beta_n/a_n \to C_1$. □

4. Examples

We now discuss some interesting examples on nonregular estimates which fall under the ambit of our results.

Example 1. Suppose that we wish to estimate the (unique) median θ_0 of a distribution $F(\cdot)$ which does not necessarily have two derivatives at the median. This situation has been considered by Huang et al. (1996) and Knight (1998a). We show how our general result can be applied to this situation. Take $f(\theta, x) = |x - \theta| - |x|$ and for $h > 0$,

$$Q(\theta + h) - Q(\theta) = 2hF(\theta + h) - h - 2E[(X - \theta)I(\theta < X \leqslant \theta + h)].$$

Table 1
Quantiles from simulated and bootstrap distributions of the sample median from $F(x) = 0.99\mathrm{N}(0,9) + 0.01\delta_0$ distribution

Quantile(%)	Simulated	B1	B2	B3
5	−2.31	−2.89	−3.50	−3.97
25	−0.49	−1.20	−1.12	−1.18
50	0.00	−0.02	−0.02	−0.01
75	0.48	0.13	0.63	0.73
95	2.26	1.86	2.47	2.78

Using this, it can be easily checked that at any point θ where F is continuous, we have $Q'(\theta) = 2F(\theta) - 1$.

Suppose that both the right and the left derivatives of F exist at θ_0. Let these be denoted by λ^+ and λ^-, respectively. Note that the usual central limit theorem for the sample median requires $\lambda^+ = \lambda^-$. Assume that this is not necessarily the case and thus F is not necessarily differentiable.

It can be checked that if we take $a_n = n^{1/2}$ and $u \geqslant 0$, then, keeping in mind that $F(\theta_0) = 1/2$,

$$n^{1/2} a_n Q(\theta_0 + u/a_n) - Q(\theta_0)$$
$$= 2un^{1/2}[F(\theta + u/a_n) - F(\theta_0)] - 2n^{1/2} a_n E[(X - \theta) I(\theta < X \leqslant \theta_0 + u/a_n)].$$

It can then be easily checked that the two terms above converge to $2\lambda^+ u^2$ and $\lambda^+ u^2$ respectively and hence the entire expression converges to $\lambda^+ u^2$. Similarly it converges to $\lambda^- u^2$ when $u < 0$.

Thus, Assumption A(i) holds with

$$\Psi(u) = \begin{cases} \lambda^+ u^2 & \text{if } u \geqslant 0, \\ \lambda^- u^2 & \text{if } u < 0. \end{cases}$$

Let us consider the mixture distribution $F(x) = 0.99\mathrm{N}(0,9) + 0.01\delta_0$, where δ_0 is the degenerate mass at 0. It is not difficult to see that for all $t \neq 0$, F is smooth, and at $t = 0$, it has unequal left and right derivatives. We take a sample of size $n = 1000$ from this distribution, and want to estimate the distribution of the sample median using resampling. Our choices of resampling techniques are

(B1) The m out of n bootstrap with $m = 150$.
(B2) The m out of n bootstrap with $m = 470$.
(B3) The weights w_{ni} are i.i.d. Gamma random variables, with shape parameter 0.1 and scale parameter 10, i.e., the mean of these weights are 1 and variance is 10.

The choice of the resampling techniques require some justification. According to Theorem 3.1 resampling techniques for which $\sigma_n^2 \to \infty$ are consistent. This however, is of little use in practice, where one must choose techniques based on the subjective criterion of "large" σ_n^2. Since for m out of n bootstrap, $\sigma_n^2 = n/m$, it would seem that choosing m smaller would produce better results. However, with smaller m the bootstrap distribution of the sample median has erratic tails, so clearly there are other factors to consider. In (B3) we have used absolutely continuous weights, and our results, presented in Table 1 suggest that they are just as good as the m out of n bootstraps.

In Table 1, we report quantiles from the simulated distribution of the sample median from F, and quantiles of the different bootstrap approximations of it.

Example 2. Consider again the median problem but now assume that $F(\cdot)$ satisfies

$$F(\theta_0 + x) - F(\theta_0) = \begin{cases} \lambda^+ |x|^\alpha L(|x|) & \text{if } x \geqslant 0, \\ -\lambda^- |x|^\alpha L(|x|) & \text{if } x < 0 \end{cases}$$

for some $\alpha > 0$ where L is a slowly varying function at 0. Then, $a_n = n^{1/(2\alpha))}L_1(n)$, where L_1 is a slowly varying function at infinity. Also, for a constant K,

$$(\nabla\Psi)(t) = \begin{cases} K\lambda^+|t|^\alpha & \text{if } t \geqslant 0, \\ -K\lambda^-|t|^\alpha & \text{if } t < 0. \end{cases}$$

Note that for $\alpha < 1$ the density $f(x)$ (if it exists) will have a singularity at $x = \theta$, and for $\alpha > 1$, $f(\theta) = 0$. Knight (1998a) considered this example also.

Example 3. Suppose F is as in Example 2, and we have the least absolute deviation regression problem, i.e., the pairs $(y_i, \mathbf{x}_i)$ are observed for $i = 1, \ldots, n$, and the regression parameter β_0 defined as $\arg\min_\beta E|y_i - \beta^T\mathbf{x}_i|$ is the parameter of interest. Knight (1998b) has example of cases where $\mathbf{x}_i$'s are nonrandom as well as cases where they are random.

Further examples can be obtained from Knight (1998b), Bose (1998) and Niemiro (1992).

Acknowledgements

The authors are grateful to the referee for her/his careful reading and constructive comments. In particular she pointed out quite a few crucial misprints and also an error in the original simulation experiment.

References

Arenal-Gutierrez, E., Matran, 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., Fukuchi, J.-I., 1997a. Bootstrapping extremes of i.i.d. random variables. Proceedings of Conference on Extreme Value Theory and its Applications. NIST, Maryland.

Athreya, K.B., Fukuchi, J.-I., 1997b. Confidence intervals for endpoints of a C.D.F. via bootstrap. J. Statist. Plann. Inference 58, 299–320.

Bickel, P.J., Gotze, F., van Zwet, W.R., 1997. Resampling fewer than n observations: gains, losses and remedies for losses. Statist. Sinica 7, 1–31.

Bose, A., 1998. Bahadur representation of M_m estimates. Ann. Statist. 26, 770–779.

Bose, A., Chatterjee, S., 2000. Generalised bootstrap for estimators of minimisers of convex functionals. Technical Report No. 5/00, Stat-Math Unit, Indian Statistical Institute, Calcutta, submitted for publication.

Castaing, C., Valadier, M., 1977. Convex Analysis and measurable Multifunctions. Lecture Notes in Mathematics, Vol. 580. Springer, New York.

Deheuvels, P., Mason, D., Shorack, G., 1993. Some results on the influence of extremes on the bootstrap. Ann. Inst. H. Poincare 29, 83–103.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1–26.

Haberman, S.J., 1989. Concavity and estimation. Ann. Statist. 17, 1631–1661.

Hjort, N.L., Pollard, D., 1993. Asymptotics for minimisers of convex processes. Statistical Research Report, University of Oslo.

Huang, J.S., Sen, P.K., Shao, J., 1996. Bootstrapping a sample quantile when the density has a jump. Statist. Sinica 6, 299–309.

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

Knight, K., 1998a. Bootstrapping sample quantiles in non-regular cases. Statist. Probab. Lett. 37, 259–267.

Knight, K., 1998b. Limit distributions for L_1 regression estimators under general conditions. Ann. Statist. 26, 755–770.

Niemiro, W., 1992. Asymptotics for M-estimators defined by convex minimization. Ann. Statist. 20 (3), 1514–1533.

Pollard, D., 1988. Asymptotics for least absolute deviation regression estimators. Technical Report, Department of Statistics, Yale University.

Praestgaard, J., Wellner, J., 1993. Exchangeably weighted bootstrap of the general empirical process. Ann. Probab. 21, 2053–2086.

Rockafellar, R.T., 1970. Convex Analysis. Princeton University Press, Princeton, NJ.

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