Generalized bootstrap for estimators of minimizers of convex functions

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

[a] *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, 700108 Kolkata, India*

[b] *Department of Mathematics and Statistics, University of Nebraska, Lincoln, USA*

Abstract

We introduce a generalized bootstrap technique for estimators obtained by minimizing functions that are convex in the parameter. We establish the consistency of these schemes via representation theorems. A number of classical resampling schemes, like the delete-d jackknife may be treated as special cases of this generalized bootstrap; and new ways of resampling are also introduced. Some of the schemes are computationally more efficient than classical techniques.

MSC: primary 62G09; secondary 62G20; 62F12; 62F40

Keywords: Bootstrap; Jackknife; M_m estimators; U-statistics; Oja median; L_1 median

1. Introduction

Let $Z_1, \ldots, Z_m$ be m independent and identically distributed (i.i.d.) copies of a Z-valued random variable and let $f(a, z)$ be a real measurable function defined for $a \in R^d$, $d \geqslant 1$, $z \in Z^m$, $m \geqslant 1$. Consider the function

$$Q(a) = E f(a, Z_1, \ldots, Z_m). \tag{1.1}$$

Assume that there is a unique $a_* \in R^d$ such that

$$Q(a_*) = \min_a Q(a), \tag{1.2}$$

a_* is the unknown parameter to be estimated from the data. Suppose that $Z_1,\ldots,Z_n$ is an i.i.d. sample. Then we may consider a sample analogue of (1.1), namely,

$$Q_n(a)=\binom{n}{m}^{-1}\sum_{1\leqslant i_1<\cdots i_m\leqslant n} f(a,Z_{i_1},Z_{i_2},\ldots,Z_{i_m}) \tag{1.3}$$

and minimize $Q_n(a)$. Let a_n be such that

$$Q(a_n)=\min_a Q_n(a). \tag{1.4}$$

The statistic a_n, which can be measurably but not necessarily uniquely selected, were introduced by Huber (1964) who called them M_m estimators. Since most criteria functions in practice are convex in the parameter, we shall assume that $f(a,z)$ is convex in a.

This class of estimators includes a large number of well-known estimators in univariate and multivariate data. The primary example is the maximum likelihood estimator when the likelihood is log-concave. Three other common examples are the mean ($m=1$, $d=1$ and $f(a,z)=(a-z)^2-z^2$); the median ($m=1$, $d=1$ and $f(a,z)=|a-z|-|z|$), and the sample variance ($m=2$, $d=2$ and $f(a,z_1,z_2)=(a-(z_1-z_2)^2/2)^2-(z_1-z_2)^4/4$). For a function $g(x_1,\ldots,x_m)$ which is symmetric in its arguments and $f(a,z_1,\ldots,z_m)=(a-g(z_1,\ldots,z_m))^2-g(z_1,\ldots,z_m)^2$, we obtain a_n to be the U-statistics with kernel g. Other examples are: several one-dimensional extensions of the median, such as the univariate Hodges–Lehmann estimators of location, the U-quantiles (Choudhury and Serfling, 1988), and the univariate location estimators of Maritz et al. (1977); multivariate extensions of the median, such as the multivariate U-quantiles (Helmers and Huskova, 1994), the Oja median (Oja, 1983), the L_1 median, the geometric quantiles (Chaudhuri, 1996), the multivariate Hodges–Lehmann versions of the univariate Hodges–Lehmann estimators; a univariate robust scale estimator of Bickel and Lehmann (1979), a regression coefficient estimator of Theil (see Hollander and Wolfe, 1973) and the least absolute deviation regression estimator in the random regressor case.

By exploiting properties of convex functions and some smoothness in expectations, asymptotic properties of these estimators may be established. Recent references in this direction include Habermann (1989), Niemiro (1992), and Bose (1998). For instance, it is known that under the two assumptions $f(a,z)$ is convex and a_* is unique, the estimator a_n converges to a_* almost surely as $n\to\infty$.

Let $g(a,z)$ be a measurable subgradient of $f(a,z)$, that is

$$f(a,z)+(b-a)^{\mathrm{T}}g(a,z)\leqslant f(b,z) \tag{1.5}$$

holds for all $a,b\in R^d$, $z\in Z^m$. Let $D(a)$ and $D^2(a)$ be, respectively, the gradient and the matrix of second derivatives of $Q(a)$ whenever they exist. Let $H=D^2(a_*)$. It may be noted that the subgradient need not be uniquely defined unless one considers it as a set-valued function. This non-uniqueness, however, does not affect the results of the present paper. The following result has been established by Habermann (1989) and Niemiro (1992) for $m=1$, and Bose (1998) for general m.

Theorem 1.1 (Habermann, 1989; Niemiro, 1992; Bose, 1998). *Under assumptions* (A.1)–(A.5) (described in Section 2)

$$n^{1/2}(a_n - a_*) = -n^{1/2}H^{-1}S_n + o_P(1), \tag{1.6}$$

where $S_n = \binom{n}{m}^{-1} \sum_{1 \leqslant i_1 < \cdots i_m \leqslant n} g(a_*, Z_{i_1}, Z_{i_2}, \ldots, Z_{i_m})$.

The asymptotic normality of M_m estimators follow immediately; in particular the asymptotic normality of all the estimators listed above under suitable conditions on the distribution of Z. The limiting dispersion matrix equals $m^2H^{-1}\Sigma H^{-1}$, where Σ is the dispersion matrix of $g_1(Z_1) = E[g(a_*, Z_1, \ldots, Z_m)|Z_1]$. The matrices H and Σ will not, in general, be easily computable and will typically involve unknown parameters.

The aim of this paper is to study a class of "weighted bootstrap" for approximating the distribution of a_n. For every $n \geqslant 1$, and every $i_1, i_2, \ldots, i_m$ distinct, $i_j \in \{1, 2, \ldots, n\}$; $j = 1, \ldots, m$, let $\{w_{n:i_1 i_2 \ldots i_m}\}$ be real-valued non-negative random variables independent of $\{Z_i\}$. These are our "bootstrap weights". We discuss conditions on these weights later on. The bootstrap equivalent of a_n will be obtained by minimizing

$$Q_{nB}(a) = \binom{n}{m}^{-1} \sum_{1 \leqslant i_1 < \cdots i_m \leqslant n} w_{n:i_1 i_2 \ldots i_m} f(a, Z_{i_1}, Z_{i_2}, \ldots, Z_{i_m}). \tag{1.7}$$

Let $\{a_{nB}\}$ be chosen in a measurable way satisfying

$$Q_{nB}(a_{nB}) = \min_a Q_{nB}(a). \tag{1.8}$$

The above scheme includes suitable generalizations of different known resampling techniques like the classical bootstrap of Efron (1979), the Bayesian bootstrap of Rubin (1981), the k out of n bootstrap and the different delete-d jackknives.

For example, if $m = 1$, Efron's classical bootstrap is obtained by using weights $(w_{n:1}, \ldots, w_{n:n}) \sim Multinomial(n; 1/n, \ldots, 1/n)$. Now suppose $m > 1$, and consider bootstrapping the U-statistics $U_n = \binom{n}{m}^{-1} \sum_{1 \leqslant i_1 < \cdots i_m \leqslant n} g(Z_{i_1}, Z_{i_2}, \ldots, Z_{i_m})$. Efron's classical bootstrap of U_n is given by

$$U_{nB} = \binom{n}{m}^{-1} \sum_{1 \leqslant i_1 < \cdots i_m \leqslant n} g(Z_{i_1}^*, Z_{i_2}^*, \ldots, Z_{i_m}^*),$$

where Z_i^* are i.i.d. from the empirical distribution of the $Z_1, Z_2, \ldots, Z_n$. In this scheme, sample values may be repeated and so kernel values like $g(X_1, X_1, \ldots)$ may appear in the bootstrap estimate. As a consequence, this bootstrap need not be consistent when the kernel is ill behaved at points $g(x, \ldots x)$. On the other hand, defining $w_{n:i_1,\ldots,i_m} = \prod_{j=1}^m w_{n:i_j}$, where $(w_{n:1}, \ldots, w_{n:n}) \sim Multinomial(n; 1/n, \ldots, 1/n)$, the bootstrapped U, according to definition (1.7) is

$$U_{nB} = \binom{n}{m}^{-1} \sum_{1 \leqslant i_1 < \cdots i_m \leqslant n} w_{n:i_1 i_2 \ldots i_m} g(Z_{i_1}, Z_{i_2}, \ldots, Z_{i_m}).$$

Now only the values of g which appear in the original U-statistics appear in the bootstrap version, thus avoiding the previous problem. This can also be termed as a "multinomial bootstrap" but is different from Efron's multinomial bootstrap.

It is also interesting to note that with $w_{n:i_1,\ldots,i_m} = \prod_{j=1}^{m} w_{n:i_j}$, and appropriate selection of $(w_{n:1},\ldots,w_{n:n})$, we obtain the delete-d jackknives for U-statistics. For $m=2$, Huskova and Janssen (1993a, b) proposed generalized bootstrap for U-statistics of degree 2, which is also part of our framework.

Our focus is to establish the theoretical consistency of a large class of resampling schemes. In Section 2, we present our main result showing that conditional on the data, a representation similar to (1.6) holds for the generalized bootstrap estimator, with the leading term being a weighted sum of exchangeable random variables. The distributional consistency can follow by applying a suitable central limit theorem for such sums. Our general consistency result immediately establishes consistency of many bootstrap procedures for a large class of estimators.

We should mention that bootstrap in some special cases of minimum contrast estimators with $m=1$ have been treated earlier. Arcones and Gine (1992) proved consistency of the multinomial bootstrap using empirical processes for some minimum contrast estimators, where they did not use convexity. Rao and Zhao (1992) have also established consistency of some generalized bootstrap estimators for M-estimators in linear regression, where the bootstrap weights are i.i.d. random variables.

In case one or more of assumption (A1)–(A6) are violated, the M_m estimators may have a non-normal limiting distribution. For $m=1$ asymptotics for such problems have been considered in Knight (1998a, b) and elsewhere; and bootstrap results have been reported in Bose and Chatterjee (2001).

A natural question is if higher order asymptotic results can be established. Such results are expected to offer theoretical comparison between the different consistent schemes. As is known, the multinomial bootstrap is second-order correct for the mean but not for the median, thereby indicating that the question is not easy to answer in general.

While it is not the goal of this article to study which of these schemes would be preferable from the computational point of view, we like to mention a few words about this. The general form of our resampling weights allows us to explore some new approaches towards resampling. Instead of the more traditional choice of the weights (a) $w_{n:i_1,\ldots,i_m} = \prod_{j=1}^{m} w_{n:i_j}$, we may consider, for example, weights of the form (b) $w_{n:i_1,\ldots,i_m} = m^{-1}\sum_{j=1}^{m} w_{n:i_j}$. Variation (b) can lead to significant computational efficiency. Note that in practice the bootstrap distribution is approximated to any degree of accuracy by B Monte Carlo steps. A small calculation shows that for B bootstrap Monte Carlo steps for a U-statistics, both the time and space complexities for method (b) is $O(n(B+1)+n^m)$ compared to $O(n^m(B+1))$ for method (a). Thus using (b) leads to economies in both computer time and memory requirements. This is not necessarily at the cost of any significant sacrifice in efficiency. We carried out a few simulation studies (see Bose and Chatterjee, 2000) with resampling the L_1 median and the Oja median in multidimensions, which show that the techniques suggested in this paper are computationally efficient compared to classical resampling techniques, but still produce very similar end results. A more comprehensive study on this aspect is underway.

2. Bootstrap asymptotic representation

All our theorems are obtained under the following set of conditions:

(A1) $f(a,z)$ is convex with respect to a for each fixed z.
(A2) The expectation in (1.1) exists and is finite for all a.
(A3) a_* satisfying (1.2) exists and is unique.
(A4) $E|g(a,Z_1,\ldots,Z_m)|^2 < \infty$ for all a in a neighborhood of a_*.
(A5) $Q(a)$ is twice differentiable at a_* and H is positive definite.
(A6) For every a, $E|a^{\mathrm{T}}(g(a_*+\varepsilon a,Z_1,\ldots,Z_m)-g(a_*,Z_1,\ldots,Z_m))|^2 \to 0$ as $\varepsilon \to 0$.

These conditions are quite mild. The second assumption may hold only for some subset of $\mathbb{R}^d$. Then all the results discussed here are valid for points in the interior of that subset. With appropriate assumptions on the distribution of Z, all the examples cited in Section 1 satisfy these conditions.

We first consider the case of $m=1$. Let $\{w_{ni},\ i=1,\ldots,n,\ n=1,2,\ldots\}$ be a triangular sequence of non-negative, *row-wise exchangeable* random variables, independent of $\{Z_1,\ldots,Z_n\}$. We use the notations P_B, E_B, V_B to, respectively, denote probabilities, expectations and variances with respect to the distribution of the weights, conditional on the given data $\{Z_1,\ldots,Z_n\}$. We henceforth drop the first suffix in the weights w_{ni} and denote it by w_i. Let $\sigma_n^2 = V_B w_i$, $W_i = \sigma_n^{-1}(w_i - 1)$. The notations k and K will be used to denote generic constants. The following conditions on the weights are assumed:

$$E_B w_1 = 1, \tag{2.1}$$

$$0 < k_1 < \sigma_n^2 = o(n), \tag{2.2}$$

$$c_{11} = \mathrm{corr}(w_i, w_j) = O(n^{-1}). \tag{2.3}$$

Remark 2.1. It may be noted that whenever $\sum_i^n w_i = C_n$ for some sequence of constants $\{C_n\}$ condition (2.3) is automatically satisfied. Thus, this condition will be automatically satisfied for the multinomial or the delete-d jackknife weights.

Theorem 2.1. *Suppose* (A1)–(A5) *hold and the row-wise exchangeable weights satisfy* (2.1)–(2.3). *Assume also that* σ_n^2/n *decreases to zero as* $n \to \infty$. *Then*

$$\sigma_n^{-1} n^{1/2}(a_{nB} - a_n) = -n^{-1/2} H^{-1} S_{nB} + r_{nB}, \tag{2.4}$$

where $S_{nB} = \sum W_i g(a_*, Z_i)$ *and* $P_B[|r_{nB}| > \varepsilon] = o_P(1)$ *for any* $\varepsilon > 0$.

The proof of Theorem 2.1 is given in the appendix.

Remark 2.2. The above theorem continues to hold if assumption $\sigma_n^2/n \downarrow 0$ as $n \to \infty$ is replaced by (A6). In most examples, (A6) is satisfied, and also $\sigma_n^2/n \downarrow 0$ is satisfied by all common resampling schemes.

We now state our bootstrap representation result for the case $m \geqslant 1$. For convenience, let us fix the notation $S = \{(i_1, \ldots, i_m) : 1 \leqslant i_1 < i_2 < \cdots < i_m \leqslant n\}$, and a typical element of S is $s = (i_1, \ldots, i_m)$. We use the notation $|s \cap t| = k$ to denote two typical subsets $s = \{i_1, \ldots, i_m\}$ and $t = \{j_1, \ldots, j_m\}$ of size m from $\{1, \ldots, n\}$, which have exactly k elements in common. We often use the same notation s for both $s = (i_1, \ldots, i_m) \in S$ and $s = \{i_1, \ldots, i_m\} \subset \{1, \ldots, n\}$. Also, let $N = \binom{n}{m}$. Let $Z_s = (Z_{i_1}, \ldots, Z_{i_m})$ when $s = (i_1, \ldots, i_m) \in S$. The notation $\sum_{|s \cap t| = j}$ denotes sum over all s and t for which $|s \cap t| = j$ holds. The bootstrap weights w_s are non-negative random variables. *We assume that for all $s \in S$, the distribution of w_s is the same.* In most applications the mean of w_s is approximately one. Keeping this in mind, define

$$\xi_n^2 = E_B(w_s - 1)^2$$

and let

$$W_s = \xi_n^{-1}(w_s - 1).$$

Assume that $E_B W_s W_t$ is a function of $|s \cap t|$ only, and let $c_j = E_B W_s W_t$ whenever $|s \cap t| = j$, for $j = 0, 1, \ldots, m$.

We assume that the following conditions are satisfied:

$$0 < k_1 < \xi_n^2 = o(n), \tag{2.5}$$

$$c_0 = O(n^{-1}), \tag{2.6}$$

$$c_j = O(1), \quad j = 1, \ldots, m-1. \tag{2.7}$$

Define $g_1(a_*, Z_1) = E[g(a_*, Z_1, \ldots, Z_m)|Z_1]$ and $f_i = \binom{n-1}{m-1}^{-1} \sum_{s: i \in s} W_s$, where the sum runs over all $s = (i_1, \ldots, i_m) \in S$ such that $I \in \{i_1, \ldots, i_m\}$.

Theorem 2.2. *Suppose* (A1)–(A5) *hold. Assume that the resampling weights satisfy* (2.5)–(2.7). *Also assume that ξ_n^2/n decreases to zero as $n \to \infty$. Then*

$$\xi_n^{-1} n^{1/2}(a_{nB} - a_n) = -n^{1/2} H^{-1} S_{nB} + r_{nB} \tag{2.8}$$

$$= -mn^{-1/2} \sum f_i H^{-1} g_1(a_*, Z_i) + R_{nB}, \tag{2.9}$$

where $S_{nB} = \binom{n}{m}^{-1} \sum_{s \in S} W_s g(a_*, Z_s)$. $P_B[|r_{nB}| > \varepsilon] = o_P(1)$ and $P_B[|R_{nB}| > \varepsilon] = o_P(1)$ for every $\varepsilon > 0$.

The proof of Theorem 2.2 is given in the appendix.

Remark 2.3. The theorem remains true if A6 replaces the bootstrap weight condition $\xi_n^2/n \downarrow 0$.

Remark 2.4. It is convenient to start with some $w_1, \ldots, w_n$, and define w_s as a function of $w_{i_1}, \ldots, w_{i_m}$. Two such functions can be easily defined, (a) $w_s = \prod_{j=1}^m w_{n:i_j}$ and (b) $w_s = m^{-1} \sum w_{n:i_j}$.

With either definition ((a) or (b)) of w_s, we still need to verify the conditions on the weights. In case of (a), this is easy when m is small but becomes increasingly cumbersome as m increases. The weights defined through (b) are easier to tackle. Interestingly, with the usual bootstrap and jackknife random variables as weights and definition (b), we have alternative forms of the usual bootstrap and jackknife schemes in case of U-functionals that have not been studied before.

To establish consistency of the generalized bootstrap distribution estimator, a bootstrap CLT is needed. In the present framework, there are several ways of establishing such a theorem, or indeed of the above asymptotic representation theorems (Theorems 1.1, 2.1 and 2.2). One way is that of the present paper, where linearization results are along the lines of Habermann (1989), Niemiro (1992), Bose (1998) and bootstrap central limit theorems are derived using arguments similar to that of Mason and Newton (1992), Praestgaard and Wellner (1993) and Arenal-Gutierrez and Matran (1996). A detailed review of this approach may be found in Bose and Chatterjee (2000). A different approach to treating stochastic convex functions is illustrated in Knight (1998a, b), Geyer (1996), Hjort and Pollard (1993). See Bose and Chatterjee (2001) for use of similar arguments in bootstrap context. A further interesting approach is present in van der Vaart and Wellner (1996, Chapters 3.2 and 3.6). The approach we use in this paper is convenient in the present context, other approaches have their own advantages too.

Our bootstrap CLT below is a special case of Lemma 4.6 of Praestgaard and Wellner (1993), which is itself a variation of Theorem 4.1 of Hajek (1961).

Theorem 2.3. *Let* $\{c_{nj};\ j=1,\ldots,n;\ n\geqslant 1\}$ *be a triangular array of constants, and let* $\{U_{nj};\ j=1,\ldots,n;\ n\geqslant 1\}$ *be a triangular array of row-exchangeable random variables such that as* $n\to\infty$.

i $$n^{-1}\sum_{j=1}^{n} c_{nj} = o(1), \tag{2.10}$$

ii $$n^{-1}\sum_{j=1}^{n} c_{nj}^2 \to \tau^2 > 0, \tag{2.11}$$

iii $$n^{-1}\max_{j=1,\ldots,n} c_{nj}^2 \to 0, \tag{2.12}$$

iv $$EU_{nj} = 0, \quad j=1,\ldots,n; \quad n\geqslant 1, \tag{2.13}$$

v $$EU_{nj}^2 = 1, \quad j=1,\ldots,n; \quad n\geqslant 1, \tag{2.14}$$

vi $$n^{-1}\sum_{j=1}^{n} U_{nj}^2 \xrightarrow{\mathrm{p}} 1, \tag{2.15}$$

vii $$\lim_{K\to\infty}\limsup_{n\to\infty}[EU_{nj}^2 I_{\{|U_{nj}|>K\}}]^{1/2} = 0. \tag{2.16}$$

Then

$$n^{-1/2}\sum_{j=1}^{n} c_{nj}U_{nj} \Rightarrow N(0,\tau^2). \tag{2.17}$$

Praestgaard and Wellner (1993) and Bose and Chatterjee (2000) have discussed sufficient conditions for the above result to hold. The variation due to Arenal-Gutierrez and Matran (1996) is also discussed in Bose and Chatterjee (2000).

We now use Theorem 2.3 to establish the consistency of the generalized bootstrap techniques in the framework of the present paper. First consider the case $m=1$. Define

$$F_n(x) = P[n^{1/2}(a_n - a_*) \leqslant x] \quad \text{and} \quad F_B(x) = P_B[\sigma_n^{-1} n^{1/2}(a_{nB} - a_n) \leqslant x],$$

where P_B is the probability conditional on the data. We have the following result:

Theorem 2.4. *Suppose $m = 1$ and assume the conditions of Theorem 2.1. Further assume that conditions of Theorem 2.3 hold with $U_{nj} = W_{nj}$. Then*

$$\sup_{x \in \mathbb{R}} |F_B(x) - F_n(x)| = o_P(1) \quad \text{as } n \to \infty.$$

Let us now consider the $m \geqslant 1$ set-up. Denote the variance of f_i by $v^2(m)$. This will, in general, be a function of m. Then the appropriate standardized bootstrap statistic is $n^{1/2}\xi_n^{-1}v_m^{-1}(a_{nB} - a_n)$. For any $c \in \mathbb{R}^d$ with $|c| = 1$, let

$$F_n(x) = \text{Prob}[n^{1/2}c^{\mathrm{T}}(a_n - a_*) \leqslant x] \quad \text{and}$$

$$F_B(x) = P_B[n^{1/2}\xi_n^{-1}v_m^{-1}(a_{nB} - a_n) \leqslant x].$$

Theorem 2.5. *Assume the conditions of Theorem 2.2. Assume that f_i's are exchangeable and satisfy (2.13)–(2.16). Further assume that $Ef_i = o(1)$. Then*

$$\sup_{x \in \mathbb{R}} |F_B(x) - F(x)| = o_P(1) \quad \text{as } n \to \infty.$$

We omit the details of the proof of Theorem 2.5.

Acknowledgements

We thank the Referees for their careful reading of the manuscript and for their constructive comments.

Appendix A. Proofs of theorems

In order to prove Theorem 2.1 we use Lemma 4 of Niemiro (1992) quoted below. In the sequel, $|a|$ stands for the Euclidean norm of a vector a.

A set $B \subset \mathbb{R}^d$ is called a *δ-triangulation* of A, if every $a \in A$ is equal to a convex combination $\sum \lambda_i b_i$ of points $b_i \in B$, such that $|b_i - a| < \delta$.

Lemma A.1 (Niemiro, 1992). *Let $A \subset A_0$ be convex subsets of $\mathbb{R}^d$ such that $|a - b| > 2\delta$ whenever $a \in A$ and $b \notin A_0$. Assume that B is a δ-triangulation of A_0.*

Suppose that h and h', are functions such that h satisfies the Lipschitz condition $|h(a)-h(b)| \leqslant L|a-b|$ for $a,b \in A_0$, and h' is convex on A_0. If

$$\sup_{b\in B} |h(b)-h'(b)| < \varepsilon$$

then

$$\sup_{a\in A} |h(a)-h'(a)| < 5L\delta + 3\varepsilon.$$

Proof of Theorem 2.1. Assume that $a_* = 0$ and $Q(a_*) = 0$. The general case when this is not necessarily true, can be reduced to this case by working with Q_1 defined by $Q_1(a) = Q(a+a_*) - Q(a_*)$.

Now define $X_{ni}(a) = [f(n^{-1/2}\sigma_n a, Z_i) - f(0,Z_i)] - n^{-1/2}\sigma_n a^{\mathrm{T}} g(0,Z_i)$ and $X_{nBi}(a) = w_{ni}X_{ni}(a)$. Then we have $E_B X_{nBi}(a) = X_{ni}(a)$ for every a and $\sum X_{nBi}(a) = nQ_{nB}(n^{-1/2}\sigma_n a) - nQ_{nB}(0) - n^{-1/2}\sigma_n a^{\mathrm{T}} S_{nW}$ where $S_{nW} = \sum w_{ni} g(0,Z_i)$. For the sake of brevity we sometimes write X_{ni} for $X_{ni}(a)$.

Let $\kappa = 3\lambda_{\min}^{-1/2}(H)$, where $\lambda_{\min}(H)$ is the least eigenvalue of H. Fix a $\delta_0 > 0$ sufficiently small such that $\kappa^2\delta_0 < 1$. Let $M > 0$ be a sufficiently large constant to be specified later on in (A.6). Let $\mathscr{A}$ be the set where both

$$\sup_{|a|\leqslant M} \sigma_n^{-2} \left| \sum_{i=1}^{n} X_{nBi}(a) - \frac{\sigma_n^2}{2} a^{\mathrm{T}} H a \right| < \delta_0, \tag{A.1}$$

$$|n^{-1/2}\sigma_n^{-1} H^{-1} S_{nW}| < M - 1 \tag{A.2}$$

hold. We will show that

$$1 - P_B[\mathscr{A}] = o_P(1). \tag{A.3}$$

On the set $\mathscr{A}$, we first show that the convex function $\Psi(a) = nQ_{nB}(n^{-1/2}\sigma_n a) - nQ_{nB}(0)$ assumes at the point $b_0 = -n^{-1/2}\sigma_n^{-1} H^{-1} S_{nW}$ a value less than its values on the sphere $|a-b_0| = \kappa\delta_0^{1/2}$.

Observe that $\Psi(a) = nQ_{nB}(n^{-1/2}\sigma_n a) - nQ_{nB}(0) = \sum_{i=1}^{n} X_{nBi}(a) + n^{-1/2}\sigma_n a^{\mathrm{T}} S_{nW}$. Also, since from (A.2) we have $|b_0| < M-1$, $b_0 \in \{a : |a| \leqslant M\}$ and hence (A.1) applies with $a = b_0$. Using (A.1) with $a = b_0$, we get $\Psi(b_0) \leqslant -2^{-1} n^{-1} S_{nW}^{\mathrm{T}} H^{-1} S_{nW} + \delta_0\sigma_n^2$.

Now since $\kappa^2\delta_0 < 1$, all a such that $|a-b_0| = \kappa\delta_0^{1/2}$ also belong to $\{a : |a| \leqslant M\}$, and hence (A.1) is applicable for such a. Write $a = b_0 + \sigma_n^{-1} g$, and it is simple algebra to see that for $a = b_0 + \sigma_n^{-1} g$ satisfying $|g| = \kappa\delta_0^{1/2}\sigma_n$, we have $\Psi(a) \geqslant -2^{-1} n^{-1} S_{nW}^{\mathrm{T}} H^{-1} S_{nW} - \delta_0\sigma_n^2 + 2^{-1} g^{\mathrm{T}} H g$. By our choice of κ, we now have that $\Psi(b_0) < \Psi(a)$ for $|a-b_0| = \kappa\delta_0^{1/2}$. This now ensures that the global minimizer of the convex function $\Psi(a)$ must lie within the sphere $|a-b_0| = \kappa\delta_0^{1/2}$.

Note that the global minimizer of $\Psi(\cdot)$ is $n^{-1/2}\sigma_n^{-1} a_{nB}$, hence we have

$$n^{1/2}\sigma_n^{-1} a_{nB} = -n^{-1/2}\sigma_n^{-1} H^{-1} S_{nW} + r_{nB3}. \tag{A.4}$$

Since δ_0 is arbitrary, and because of (A.3) we have $P_B[|r_{nB3}| > \delta] = o_P(1)$ for any $\delta > 0$.

Now divide (1.6) by σ_n and subtract it from (A.4) to get (2.4). This step again uses the lower bound condition of (2.2).

Thus, the proof is complete once (A.3) is established. For this, we show that

$$\text{For any } M > 0;\ P_B\left[\sup_{|a|\leqslant M} \sigma_n^{-2}\left|\sum X_{nBi}(a) - \sigma_n^2 a^{\mathrm{T}} Ha/2\right| > \delta_0\right] = o_P(1), \tag{A.5}$$

$$\exists M > 0 \text{ such that } P_B[|\sigma_n^{-1} n^{-1/2} H^{-1} S_{nW}| \geqslant M] = o_P(1). \tag{A.6}$$

Proof of (A.5). Fix an $M > 0$. For fixed $\delta_0 > 0$, get $\delta > 0$ and $\varepsilon > 0$ such that $M_1 = M + 2\delta$ and $\delta_0 > 5M_1\lambda_{\max}(H)\delta + 3\varepsilon$, where $\lambda_{\max}(H)$ is the maximum eigenvalue of H. Consider the set $A_1 = \{a : |a| \leqslant M_1\}$ and let $B = \{b_1, \ldots, b_T\}$ be a finite δ-triangulation of A_1. Note that with $A = \{a : |a| \leqslant M\}$, $L = M_1\lambda_{\max}(H)$, $h(a) = a^{\mathrm{T}} Ha/2$ and $h'(a) = \sigma_n^{-2}\sum w_i X_{ni}(a)$, all the conditions of Lemma A.1 are satisfied. Now

$$\begin{aligned}
&P_B\left[\sup_{|a|\leqslant M} \sigma_n^{-2}\left|\sum X_{nBi}(a) - \sigma_n^2 a^{\mathrm{T}} Ha/2\right| > \delta_0\right] \\
&\leqslant P_B\left[\sup_{|a|\leqslant M} \sigma_n^{-2}\left|\sum X_{nBi}(a) - \sigma_n^2 a^{\mathrm{T}} Ha/2\right| > 5L\delta + 3\varepsilon\right] \\
&\leqslant P_B\left[\sup_{a\in B} \sigma_n^{-2}\left|\sum X_{nBi}(a) - \sigma_n^2 a^{\mathrm{T}} Ha/2\right| > \varepsilon\right] \qquad \text{(A.7)} \\
&\leqslant \sum_{j=1}^{T} P_B\left[\sigma_n^{-2}\left|\sum X_{nBi}(b_j) - \sigma_n^2 b_j^{\mathrm{T}} Hb_j/2\right| > \varepsilon\right] \\
&\leqslant \sum_{j=1}^{T} P_B\left[\sigma_n^{-1}\left|\sum W_i X_{ni}(b_j)\right| > \varepsilon/2\right] + \sum_{j=1}^{T} I_{\{\sigma_n^{-2}|\sum X_{ni}(b_j) - \sigma_n^2 b_j^{\mathrm{T}} Hb_j/2| > \varepsilon/2\}} \qquad \text{(A.8)} \\
&\leqslant \sigma_n^{-2}\sum_{j=1}^{T} k\sum_{j=1}^{n} X_{ni}^2(b_j) + \sum_{j=1}^{T} I_{\{\sigma_n^{-2}|\sum X_{ni}(b_j) - \sigma_n^2 b_j^{\mathrm{T}} Hb_j/2| > \varepsilon/2\}} \qquad \text{(A.9)} \\
&= o_P(1). \qquad \text{(A.10)}
\end{aligned}$$

In the above calculations, (A.7) follows from Lemmas A.1. (A.9) follows from (A.7) by the following argument:

$$\begin{aligned}
&P_B\left[\sigma_n^{-1}\left|\sum_i W_i X_{ni}(b_j)\right| > \varepsilon/2\right] \\
&\leqslant 4\varepsilon^{-2}\sigma_n^{-2} E_B\left[\sum_i W_i X_{ni}(b_j)\right]^2
\end{aligned}$$

$$
\begin{aligned}
&= 4\varepsilon^{-2}\sigma_n^{-2}\left[\sum_i X_{ni}^2(b_j) + c_{11}\sum_{i\neq k} X_{ni}(b_j)X_{nk}(b_j)\right]\\
&= 4\varepsilon^{-2}\sigma_n^{-2}\left[(1-c_{11})\sum_i X_{ni}^2(b_j) + c_{11}\left(\sum_i X_{ni}(b_j)\right)^2\right]\\
&\leqslant 4\varepsilon^{-2}\sigma_n^{-2}\left[(1-c_{11})\sum_i X_{ni}^2(b_j) + nc_{11}\left(\sum_i X_{ni}^2(b_j)\right)\right]\\
&\leqslant C\sum_i X_{ni}^2(b_j).
\end{aligned}
$$

To obtain (A.10) from (A.9) observe that the number of terms in the sum over j is finite. We first show that each term in the sum is $o_P(1)$.

In order to prove $\sigma_n^{-2}\sum_i X_{ni}^2(b) = o_P(1)$, we will show that $\sigma_n^{-2}\sum_i EX_{ni}^2(b) = n\sigma_n^{-2}EX_{n1}^2(b) = o(1)$. Note that for every a, we have

$$
\begin{aligned}
&n\sigma_n^{-2}E[f(n^{-1/2}\sigma_n a, Z_1) - f(0,Z_1) - n^{-1/2}\sigma_n a^{\mathrm{T}} g(0,Z_1)]^2\\
&\quad\leqslant E[a^{\mathrm{T}}(g(n^{-1/2}\sigma_n a, Z_1) - g(0,Z_1))]^2
\end{aligned}
$$

from (4.4) of Niemiro (1992). By using (A6), this converges to 0.

In place of (A6), if we had assumed that $n^{-1/2}\sigma_n \downarrow 0$, then the above again follows by using arguments of Niemiro (1992, p. 1522).

This establishes that for fixed b, $\sigma_n^{-2}\sum_i X_{ni}^2(b) = o_P(1)$.

For the proof of $\sigma_n^{-2}[\sum_i X_{ni}(b) - \sigma_n^2 b^{\mathrm{T}}Hb/2] = o_P(1)$, we write

$$
\begin{aligned}
&\sigma_n^{-2}\left[\sum_i X_{ni}(b) - \sigma_n^2 b^{\mathrm{T}}Hb/2\right]\\
&\quad=\sigma_n^{-2}\left[\sum_i X_{ni}(b) - E\sum_i X_{ni}(b) + E\sum_i X_{ni}(b) - \sigma_n^2 b^{\mathrm{T}}Hb/2\right].
\end{aligned}
$$

Note that $E\sum_i X_{ni}(b) = nQ(n^{-1/2}\sigma_n b)$, for which we have the Taylor series expansion $Q(a) = a^{\mathrm{T}}Ha/2 + o(|a|^2)$ for a near 0, since by our choice the value of $Q(\cdot)$ and its derivative at 0 are 0. This shows that $n\sigma_n^{-2}Q(n^{-1/2}\sigma_n b) - b^{\mathrm{T}}Hb/2 = o(1)$.

In order to prove $\sigma_n^{-2}[\sum_i X_{ni}(b) - E\sum_i X_{ni}(b)] = o_P(1)$ we have

$$
\begin{aligned}
&\sigma_n^{-4}E\left[\sum_i X_{ni}(b) - E\sum_i X_{ni}(b)\right]^2\\
&\quad\leqslant k_1^{-2}\sigma_n^{-2}E\left[\sum_i X_{ni}(b) - E\sum_i X_{ni}(b)\right]^2
\end{aligned}
$$

$$= k_1^{-2}\sigma_n^{-2}E\left[\sum_i (X_{ni}(b) - EX_{ni}(b))\right]^2$$

$$= k_1^{-2}\sigma_n^{-2}E\sum_i (X_{ni}(b) - EX_{ni}(b))^2$$

$$\leqslant nk_1^{-2}\sigma_n^{-2}E(X_{n1}(b))^2$$

$$= o(1).$$

This proves (A.5). □

Proof of (A.6). Let λ_1 be the minimum eigenvalue of H. Then

$$P_B[|\sigma_n^{-1}n^{-1/2}H^{-1}S_{nW}| > M]$$

$$\leqslant P_B[|\sigma_n^{-1}n^{-1/2}S_{nW}| > M\lambda_1]$$

$$\leqslant \frac{1}{\lambda_1^2 M^2 n\sigma_n^2} E_B\left|\sum_{i=1}^n w_i g(0,Z_i)\right|^2$$

$$\leqslant \frac{2}{\lambda_1^2 M^2 n\sigma_n^2}\left[\sigma_n^2 E_B\left|\sum_{i=1}^n W_i g(0,Z_i)\right|^2 + \left|\sum_{i=1}^n g(0,Z_i)\right|^2\right]$$

$$\leqslant \frac{2}{\lambda_1^2 M^2 n}\sum_{i,j=1}^n E_B W_i W_j g(0,Z_i)^{\mathrm{T}} g(0,Z_j) + \frac{2}{\lambda_1^2 M^2 n\sigma_n^2}\left|\sum_{i=1}^n g(0,Z_i)\right|^2$$

$$\leqslant \frac{2}{\lambda_1^2 M^2 n}\left[(1-c_{11})\sum_{i=1}^n |g(0,Z_i)|^2 + c_{11}\sum_{i,j=1}^n g(0,Z_i)^{\mathrm{T}} g(0,Z_j)\right]$$

$$+ \frac{2}{\lambda_1^2 M^2 k_1 n}\left|\sum_{i=1}^n g(0,Z_i)\right|^2$$

$$\leqslant \frac{2}{\lambda_1^2 M^2 n}\left[(1-c_{11})\sum_{i=1}^n |g(0,Z_i)|^2 + c_{11}\left|\sum_{i=1}^n g(0,Z_i)\right|^2\right]$$

$$+ \frac{2}{\lambda_1^2 M^2 k_1 n}\left|\sum_{i=1}^n g(0,Z_i)\right|^2$$

$$\leqslant \frac{K_1}{M^2 n}\sum_{i=1}^n |g(0,Z_i)|^2 + \frac{K_2}{M^2 n}\left|\sum_{i=1}^n g(0,Z_i)\right|^2$$

$$= U_M \quad \text{say}.$$

Now fix any two constants $\delta_1, \delta_2 > 0$. By choosing M large enough, we have

$$\begin{aligned}&\text{Prob}[P_B[\sigma_n^{-1}|n^{-1/2}H^{-1}S_{nW}| < M] > \delta_1]\\ &\leqslant \text{Prob}[\delta_1 < U_M]\\ &< \delta_2. \end{aligned} \tag{A.11}$$

This proves (A.6). □

Proof of Theorem 2.2. Let us establish (2.8) first. This proof is similar to that of Theorem 2.1. Define $X_{ns}(a) = f(n^{-1/2}\xi_n a, Z_s) - f(0, Z_s) - n^{-1/2}\xi_n a^{\mathrm{T}} g(0, Z_s)$, and let $X_{nBs}(a) = w_s X_{ns}(a)$, $S_{nW} = N^{-1}\sum_s w_s g(0, Z_s)$. We have $nQ_{nB}(n^{-1/2}\xi_n a) - nQ_{nB}(0) = n^{1/2}\xi_n a^{\mathrm{T}} S_{nW} + nN^{-1}\sum X_{nBs}(a)$.

By carefully following the arguments for Theorem 2.1, we only have to show that

$$\begin{aligned}&\text{For any } M > 0; P_B\left[\sup_{|a|\leqslant M} \xi_n^{-2}\left|nN^{-1}\sum X_{nBs}(a) - \xi_n^2 a^{\mathrm{T}} Ha/2\right| > \delta_0\right]\\ &= o_P(1), \end{aligned} \tag{A.12}$$

$$\exists M > 0 \text{ such that } P_B[|\xi_n^{-1} n^{1/2} H^{-1} S_{nW}| \geqslant M] = o_P(1) \tag{A.13}$$

for a sufficiently small constant $\delta_0 > 0$. The proof of these are similar to that of (A.5) and (A.6).

Using the facts $n^{1/2}N^{-1}\sum_s g(0, Z_s) = O_P(1)$ and $nN^{-2}\sum_s |g(0, Z_s)| = O_P(1)$, after some algebra (A.13) is established, the algebra being similar to the corresponding calculation in Theorem 2.1.

For (A.12), again by some algebra similar to Theorem 2.1, the result is established once the following two are proved for any fixed a with $|a| \leqslant M$ and any $\varepsilon > 0$:

$$P_B\left[\xi_n^{-1} nN^{-1}\left|\sum_s W_s X_{ns}(a)\right| > \varepsilon\right] = o_P(1), \tag{A.14}$$

$$I_{\{\xi_n^{-2}|nN^{-1}\sum X_{ns}(a) - \xi_n^2 a^{\mathrm{T}} Ha/2| > \varepsilon\}} = o_P(1). \tag{A.15}$$

The proof of (A.15) follows from considering the two parts $\{\xi_n^{-2} nN^{-1}|\sum X_{ns}(a) - E\sum X_{ns}(a)| > \varepsilon\}$ and $\{\xi_n^{-2}|nN^{-1}E\sum X_{ns}(a) - \xi_n^2 a^{\mathrm{T}} Ha/2| > \varepsilon\}$ separately, and using Chebyshev's inequality for the first part and a Taylor series expansion for the second part.

For (A.14), we have

$$\begin{aligned}&P_B\left[\xi_n^{-1} nN^{-1}\left|\sum_s W_s X_{ns}(a)\right| > \varepsilon\right]\\ &\leqslant \varepsilon^{-2}\xi_n^{-2} n^2 N^{-2} E_B\left|\sum W_s X_{ns}(a)\right|^2\end{aligned}$$

$$= \varepsilon^{-2}\xi_n^{-2}n^2N^{-2}\sum_{j=0}^{m} c_j \sum_{\{s,t:|s\cap t|=j\}} X_{ns}X_{nt}$$

$$= \varepsilon^{-2}\xi_n^{-2}n^2N^{-2}\left[\sum_s |X_{ns}|^2 + \sum_{j=0}^{m-1} c_j \sum_{\{s,t:|s\cap t|=j\}} X_{ns}X_{nt}\right].$$

Note that since g is the subgradient function, we have

$$0 \leqslant X_{ns}(a) \leqslant n^{-1/2}\xi_n a^{\mathrm{T}}(g(n^{-1/2}\xi_n a, Z_s) - g(0, Z_s)).$$

By arguments similar to those of Theorem 2.1, we have $\xi_n^{-2}n^2N^{-2}\sum_s |X_{ns}|^2 = o_P(1)$. We also need to show $\xi_n^{-2}n^2N^{-2}c_j \sum_{\{s,t:|s\cap t|=j\}} X_{ns}X_{nt} = o_P(1)$ for $j=0,\ldots,m-1$. From the above argument it follows that this is established if $c_j nN^{-2}\sum_{\{s,t:|s\cap t|=j\}} 1 = O(1)$ for $j = 0,\ldots,m-1$. Note that $\sum_{\{s,t:|s\cap t|=j\}} 1 = \binom{n}{j}\binom{n-j}{2(m-j)}\binom{2(m-j)}{m-j} = O(n^{2m-j})$. The result now follows from the conditions assumed on c_j's.

In order to get the representation (2.9), we have to show

$$n^{1/2}N^{-1}\sum W_s g(0, Z_s) = mn^{-1/2}\sum_i f_i g_1(0, Z_i) + R_{nB1},$$

where for any $\delta > 0$, we have $P_B[|R_{nB1}| > \delta] = o_P(1)$.

Let $h(0, Z_s) = g(0, Z_s) - \sum_{j=1}^{m} g_1(0, Z_{i_j})$. This is a kernel of a first-order degenerate U-statistics. Using this we now have $W_s g(0, Z_s) = W_s \sum_{j=1}^{m} g_1(0, Z_{i_j}) + W_s h(0, Z_s)$. Also note that $\sum_s \sum_{j=1}^{m} W_s g_1(0, Z_{i_j}) = \binom{n-1}{m-1}\sum_{i=1}^{n} f_i g_1(0, Z_i)$. Finally, note that $E(|N^{-1}\sum h(0, Z_s)|)^{-2} = O(n^{-2})$. Now using this result and (2.7), after some algebra we obtain $P_B[|n^{1/2}N^{-1}\sum W_s h(0, Z_s)| > \delta] = o_P(1)$, and this yields (2.9). □

Proof of Theorem 2.4. First assume that the parameter space is one dimensional. The first term in the right-hand side of the bootstrap representation (2.4) can be identified in the framework of Theorem 2.3 with $c_{nj} = H^{-1}g(a_*, Z_j)$. Note that since a_* is the minimizer, we have $E(g(a_*, Z_1))=0$. Thus $\bar{a}_n$ converges to zero almost surely. Standard arguments show that assumptions (2.11) and (2.12) are satisfied almost surely under (A4). Thus by applying the above result, the asymptotic (conditional) normality of $\sigma_n^{-1}n^{1/2}(a_{nB} - a_n)$ follows since the weights $U_{nj} = W_{nj} = \sigma_n^{-1}(w_{nj} - 1)$ satisfy conditions (2.13)–(2.16). It is also easy to see that the limiting variance is $\tau^2 = H^{-2}V(g(a_*, Z_1))$ which is exactly equal to the limiting variance of $n^{1/2}(a_n - a_*)$. This establishes the consistency of the generalized bootstrap if the parameter space is one dimensional.

Then $\{c_{nj}\}$ are vector valued, but using the standard Cramer–Wold device the above argument can be easily repeated. □

References

Arcones, M.A., Gine, E., 1992. On the bootstrap of M-estimators and other statistical functionals. In: LePage, R., Billard, L. (Eds.), Exploring the Limits of Bootstrap. Wiley, New York, pp. 13–48.

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.

Bickel, P.J., Lehmann, E.L., 1979. Descriptive statistics for nonparametric models, IV. Spread. In: Jureckova, J. (Ed.), Contributions to Statistics. Academia, Prague, pp. 33–40.

Bose, A., 1998. Bahadur representation of M_m estimates. Ann. Statist. 26, 771–777.

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. Available at www.math.unl.edu/~schatterjee.

Bose, A., Chatterjee, S., 2001. Generalised bootstrap in non-regular M-estimation problems. Statist. Probab. Lett. 55, 319–328.

Chaudhuri, P., 1996. On a geometric notion of quantiles for multivariate data. J. Amer. Statist. Assoc. 91, 862–872.

Choudhury, J., Serfling, R.J., 1988. Generalized order statistics, Bahadur representations, and sequential nonparametric fixed width confidence intervals. J. Statist. Plann. Inference 19, 269–282.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 101–118.

Geyer, C.J., 1996. On the asymptotics of convex stochastic optimization. Unpublished manuscript.

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

Hajek, J., 1961. Some extensions of the Wald–Wolfowitz–Noether theorem. Ann. Math. Statist. 32, 506–523.

Helmers, R., Huskova, M., 1994. Bootstrapping multivariate U quantiles and related statistics. J. Multivariate Anal. 49 (1), 97–109.

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

Hollander, M., Wolfe, D.A., 1973. Nonparametrical Statistical Methods. Wiley, New York.

Huber, P.J., 1964. Robust estimation of a location parameter. Ann. Math. Statist. 35, 73–101.

Huskova, M., Janssen, P., 1993a. Generalized bootstrap for studentized U-statistics: a rank statistics approach. Statist. Probab. Lett. 16, 225–233.

Huskova, M., Janssen, P., 1993b. Consistency of generalized bootstrap for degenerate U-statistics. Ann. Statist. 21, 1811–1823.

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.

Maritz, J.S., Wu, M., Staudte, R.G., 1977. A location estimator based on U statistic. Ann. Statist. 5, 779–786.

Mason, D.A., Newton, M.A., 1992. A rank statistics approach to the consistency of a general bootstrap. Ann. Statist. 20, 1611–1624.

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

Oja, H., 1983. Descriptive statistics for multivariate distribution. Statist. Probab. Lett. 1, 327–333.

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

Rao, C.R., Zhao, L.C., 1992. Approximations to the distribution of M-estimates in linear models by randomly weighted bootstrap. Sankhyā Ser. A 54, 323–331.

Rubin, D.B., 1981. The Bayesian bootstrap. Ann. Statist. 9, 130–134.

van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes with Applications to Statistics. Springer Series in Statistics. Springer, New York.