MULTIVARIATE L_1-NORM ESTIMATION AND THE VULNERABLE BOOTSTRAP

by

Pranab Kumar Sen

Department of Biostatistics, University of
North Carolina at Chapel Hill, NC.

Institute of Statistics Mimeo Series No. 1895

March 1992

MULTIVARIATE L_1-NORM ESTIMATION AND THE VULNERABLE BOOTSTRAP

Pranab Kumar Sen

Rotation-invariant L_1-norm estimation of location of a multivariate distribution is considered. Although this method rests on bounded score functions in a multivariate M-estimation setup, its robustness properties depend on the distribution of the reciprocal distances of the sample observations as well as on interdependence of the coordinate variates. In this context, the jackknife, bootstrap and some other resampling methods can be adapted to estimate the asymptotic dispersion matrix, but these are generally biased. A comparative picture of jackknifing and bootstrapping is presented.

1. L_1-NORM ESTIMATOR

Let $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ be n independent and identically distributed random vectors (i.i.d.r.v.) with a distribution function (d.f.) $F_{\underset{\sim}{\theta}}(\underset{\sim}{x}) = F(\underset{\sim}{x} - \underset{\sim}{\theta})$, $\underset{\sim}{x} \in R^p$, $\underset{\sim}{\theta} \in \Theta \subseteq R^p$, where the form of F does not depend on the location parameter $\underset{\sim}{\theta}$. For a nonnegative definite (nnd) matrix $\underset{\sim}{Q}$, consider the quadratic norm

$$\| \underset{\sim}{d} \|^2_{\underset{\sim}{Q}} = \underset{\sim}{d}' \underset{\sim}{Q} \underset{\sim}{d} \quad \text{and let} \quad \| \underset{\sim}{d} \| = \| \underset{\sim}{d} \|_{\underset{\sim}{I}} . \tag{1.1}$$

For simplicity, we work with $\underset{\sim}{Q} = \underset{\sim}{I}$, and define

$$\rho_n(\underset{\sim}{\theta}) = \Sigma_{i=1}^n \| \underset{\sim}{X}_1 - \underset{\sim}{\theta} \| , \quad \underset{\sim}{\theta} \in \Theta . \tag{1.2}$$

The L_1-norm estimator $\hat{\underset{\sim}{\theta}}_n$ of $\underset{\sim}{\theta}$ is defined by

$$\hat{\underset{\sim}{\theta}}_n = \text{Arg. min}\{\rho_n(\underset{\sim}{\theta}) : \ \underset{\sim}{\theta} \in R^p\} . \tag{1.3}$$

For $p = 1$, minimization of (1.2) with respect to θ leads to the estimator

$$\hat{\underset{\sim}{\theta}}_n = \text{median}(X_1, \ldots, X_n) ,$$

although the usual differentiation technique does not apply here. For $p \geq 2$, $\rho_n(\underset{\sim}{\theta})$ is differentiable with respect to $\underset{\sim}{\theta}$, and the minimization process yields the following estimating equation:

$$\Sigma_{i=1}^n \| \underset{\sim}{X}_i - \hat{\underset{\sim}{\theta}}_n \|^{-1} (\underset{\sim}{X}_i - \hat{\underset{\sim}{\theta}}_n) = \underset{\sim}{0} , \tag{1.4}$$

where conventionally, we let $\| \underset{\sim}{0} \|^{-1} \underset{\sim}{0} = \underset{\sim}{0}$, as is needed to eliminate the indeterminacy in (1.4) if $\hat{\underset{\sim}{\theta}}_n$ coincides with one of the observations. Generally, an iterative procedure is needed to solve for $\hat{\underset{\sim}{\theta}}_n$ in (1.4). It may be noted that (1.4) represents an estimating equation closely related to that in multivariate M-estimation of location (viz., Singer and Sen [7]), although the associated score function involves weights which depend in a rather involved manner on the estimator $\hat{\underset{\sim}{\theta}}_n$ (identifiable as the center of gravity). If we consider the group of transformations: $\underset{\sim}{X} \rightarrow \underset{\sim}{X}^* = \underset{\sim}{A}\underset{\sim}{X}$ and $\underset{\sim}{\theta} \rightarrow \underset{\sim}{\theta}^* = \underset{\sim}{A}\underset{\sim}{\theta}$ where $\underset{\sim}{A}$ is an orthogonal matrix (i.e., $\underset{\sim}{A}'\underset{\sim}{A} = \underset{\sim}{I}_p$), then (1.4) yields rotation-equivariance of $\hat{\underset{\sim}{\theta}}_n$. However, $\hat{\underset{\sim}{\theta}}_n$

may not be generally equivariant under affine transformations: $\underset{\sim}{X} \to \underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{X}$, $\underset{\sim}{\theta} \to \underset{\sim}{a} + \underset{\sim}{B}\underset{\sim}{\theta}$ where $\underset{\sim}{B}$ is nnd. In passing, we may remark that instead of the simple location model, we could have considered a general multivariate linear model:

$$\underset{\sim}{Y}_i = \underset{\sim}{\beta}\underset{\sim}{t}_i + \underset{\sim}{e}_i , \quad i = 1, \ldots, n , \tag{1.5}$$

where the $\underset{\sim}{e}_i$ are i.i.d.r.v.'s with a d.f. F (defined on R^p) whose form does not depend on the unknown parameter $\underset{\sim}{\beta}$ (p × q matrix), and the $\underset{\sim}{t}_i$ are given q-vectors ($q \geq 1$). Parallel to (1.2), we could have taken the L_1-norm

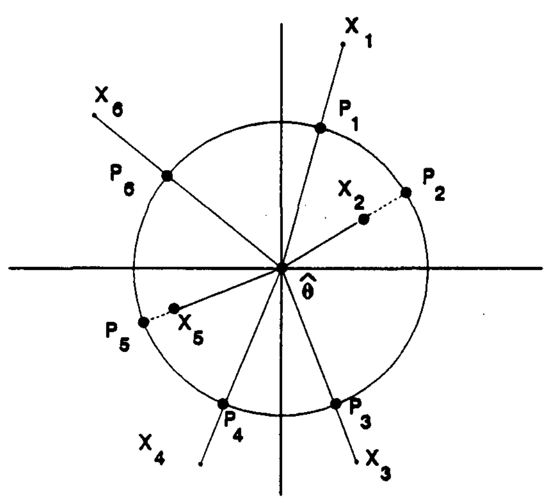
$$D_{n1}(\underset{\sim}{\beta}) = \Sigma_{i=1}^{n} \| \underset{\sim}{Y}_i - \underset{\sim}{\beta}\underset{\sim}{t}_i \| , \quad \underset{\sim}{\beta} \in R^{p \times q} , \tag{1.6}$$

and noting that $D_{n1}(\underset{\sim}{\beta})$ is convex in $\underset{\sim}{\beta}$, minimization of (1.6) with respect to $\underset{\sim}{\beta}$ would have led us to the L_1-norm estimator of β:

$$\hat{\underset{\sim}{\beta}}_n = \text{Arg. min}\{D_{n1}(\underset{\sim}{\beta}): \underset{\sim}{\beta} \in R^{p \times q}\} . \tag{1.7}$$

For the sake of simplicity of presentation, in the sequel, we consider only the location model, although the results to follow pertain to the general regression model as well. In this context, we shall restrict ourselves to $p \geq 2$, so that $\rho_n(\underset{\sim}{\theta})$ in (1.2) is differentiable with respect to $\underset{\sim}{\theta}$. For $p = 1$, the results pertaining to (1.3) are all extensively studied in the literature, and hence, we do not repeat them here.

We conclude this section with a remark on the computation of $\hat{\underset{\sim}{\theta}}_n$ in (1.4). Consider a sphere $S(\underset{\sim}{a})$ of unit radius and center at $\underset{\sim}{a}$ ($\in R^p$). Join each point $\underset{\sim}{X}_i$ to $\underset{\sim}{a}$ and obtain the point of intersection (or projection) of the line joining $\underset{\sim}{X}_i$ and $\underset{\sim}{a}$ on the sphere $S(\underset{\sim}{a})$. In this manner, the set of n points $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ are mapped into a set of n points on the spherical surface of $S(\underset{\sim}{a})$. Then $\hat{\underset{\sim}{\theta}}_n$ relates to that particular choice of $\underset{\sim}{a}$ for which the projections of these n points on the spherical surface of $S(\hat{\underset{\sim}{\theta}}_n)$ along any direction has mean zero. The following figure relates to a simple bivariate situation (with $n = 6$):



Each $\underset{\sim}{X}_i$ continues to have a bounded influence in (1.4), although not solely on a coordinatewise basis. In this respect, the situation is different from the usual multivariate M-estimation of location, where coordinatewise influence functions dominate the scenario. Thus, the L_1-norm estimator may entail a lot of computational manipulations, albeit some iterative procedures work out well. The estimator is, however, highly non-linear.

2. ASYMPTOTICS VIA LINEARITY APPROACH

Keeping in mind the estimating equations in (1.4), we define the following stochastic processes:

$$\underset{\sim}{M}_n(\underset{\sim}{t}) = n^{-1/2} \Sigma_{i=1}^n \| \underset{\sim}{X}_i - \underset{\sim}{\theta} - \underset{\sim}{t} \|^{-1} (\underset{\sim}{X}_i - \underset{\sim}{\theta} - \underset{\sim}{t}), \quad \underset{\sim}{t} \in R^p . \tag{2.1}$$

Let then $\underset{\sim}{Z}_n = \{\underset{\sim}{Z}_n(\underset{\sim}{u}), \underset{\sim}{u} \in R^p\}$ be defined by

$$\underset{\sim}{Z}_n(\underset{\sim}{u}) = \{\underset{\sim}{M}_n(n^{-1/2}\underset{\sim}{u}) - \underset{\sim}{M}_n(\underset{\sim}{0}) + \underset{\sim}{V}_n\underset{\sim}{u}\} , \quad \underset{\sim}{u} \in R^p , \tag{2.2}$$

where

$$\underset{\sim}{V}_n = n^{-1} \Sigma_i^n \| \underset{\sim}{X}_i - \underset{\sim}{\theta} \|^{-1} \{\underset{\sim}{I}_p - \| \underset{\sim}{X}_i - \underset{\sim}{\theta} \|^{-2} (\underset{\sim}{X}_i - \underset{\sim}{\theta})(\underset{\sim}{X}_i - \underset{\sim}{\theta})'\} . \tag{2.3}$$

Then, we have the following.

Theorem 2.1. Assume that there exists a $\delta > 0$, such that

$$\nu_F(\underset{\sim}{\xi}) = E_F \| \underset{\sim}{X} - \underset{\sim}{\xi} \|^{-1} < \infty , \quad \text{uniformly in } \underset{\sim}{\xi}: \| \underset{\sim}{\xi} - \underset{\sim}{\theta} \| < \delta . \tag{2.4}$$

Then, for every K: $0 < K < \infty$, as $n \to \infty$,

$$\sup\{ \| \underset{\sim}{Z}_n(\underset{\sim}{u}) \| : \| \underset{\sim}{u} \| < K\} \to 0 , \quad \text{in probability.} \tag{2.5}$$

The proof of (2.5) follows along the lines of Jurečková and Sen [4], and hence, the details are omitted. In passing, we may note that for $p = 1$, V_n is 0, so that the version in (2.2) and (2.5) is not true, but a more refined result (along the line of the classical Bahadur representation of sample quantiles) holds (viz., Jurečková and Sen [3]). To make use of Theorem 2.1 in the context of the asymptotic distribution theory of $\hat{\underset{\sim}{\theta}}_n$, let us define

$$\begin{aligned} \underset{\sim}{\Delta}_F &= E\underset{\sim}{V}_n \\ &= E_F\{ \| \underset{\sim}{X} - \underset{\sim}{\theta} \|^{-1} [\underset{\sim}{I}_p - \| \underset{\sim}{X} - \underset{\sim}{\theta} \|^{-2}(\underset{\sim}{X} - \underset{\sim}{\theta})(\underset{\sim}{X} - \underset{\sim}{\theta})'\} , \end{aligned} \tag{2.6}$$

$$\Sigma_F = E_F\{ \| \underset{\sim}{X} - \underset{\sim}{\theta} \|^{-2} (\underset{\sim}{X} - \underset{\sim}{\theta})(\underset{\sim}{X} - \underset{\sim}{\theta})'\} \tag{2.7}$$

and

$$\underset{\sim}{\Gamma}_F \; \underset{\sim}{\Delta}_F^{-1} \, \underset{\sim}{\Sigma}_F \, \underset{\sim}{\Delta}_F^{-1} , \tag{2.8}$$

where we assume that Δ_F and Σ_F are both nnd, so that Γ_F is also nnd. Note that by the Khintchine strong law of large numbers, whenever $E_F \| X - \theta \|^{-1} < \infty$, as $n \to \infty$,

$$V_n \to \Delta_F \text{ almost surely (a.s.)} \tag{2.9}$$

Also, by (2.1) and the multivariate central limit theorem, as $n \to \infty$,

$$M_n(0) \xrightarrow{\mathfrak{D}} \mathcal{N}_p(0, \Sigma_F) . \tag{2.10}$$

Therefore, from (1.4), (2.2), (2.5) and (2.10), we obtain that as $n \to \infty$,

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathfrak{D}} \mathcal{N}_p(0, \Gamma_F) , \tag{2.11}$$

which may be used to draw statistical conclusions on θ. For an alternative approach to the study of (2.11), we may refer to Bai et al. [1] where other pertinent references are also cited.

3. STATISTICAL INFERENCE BASED ON L_1-NORM ESTIMATORS

In an asymptotic setup, in order to construct a confidence region for θ or to test for a null hypothesis on θ, a valid use of (2.11) rests on a consistent estimator of Γ_F . Looking at (2.6), (2.7) and (2.8), we may propose the estimator

$$\hat{\Gamma}_n = \hat{\Delta}_n^{-1} \hat{\Sigma}_n \hat{\Delta}_n^{-1} , \tag{3.1}$$

where

$$\hat{\Delta}_n = \frac{1}{n} \sum_{i=1}^{n} \| X_i - \hat{\theta}_n \|^{-1} \{ I_p - \| X_i - \hat{\theta}_n \|^{-2} (X_i - \hat{\theta}_n)(X_i - \hat{\theta}_n)' \} \tag{3.2}$$

and

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^{n} \| X_i - \hat{\theta}_n \|^{-2} (X_i - \hat{\theta}_n)(X_i - \hat{\theta}_n)' . \tag{3.3}$$

It seems that to establish the consistency of $\hat{\Delta}_n$ and $\hat{\Sigma}_n$ (and hence, $\hat{\Gamma}_n$), we may need some extra regularity condition. One possibility is to assume that

$$E_\theta \Big\{ \sup \| \frac{1}{\| X - \theta - t \|} \{ I - \| X - \theta - t \|^{-2} (X - \theta - t)(X - \theta - t)' \}$$

$$- \{ I - \| X - \theta \|^{-2} ((X - \theta)(X - \theta)' \} \| : \ \| t \| < \delta \Big\} \to 0 \ \text{ as } \ \delta \downarrow 0 , \tag{3.4}$$

so that the consistency of $\hat{\theta}_n$ and (3.4) along with the Khintchine strong law of large numbers would imply the consistency of $\hat{\Gamma}_n$. Since the summands in (3.3) are points on the unit sphere $S(\hat{\theta}_n)$ (and the convention after (1.4) is adopted), for $\hat{\Sigma}_n$, the regularity conditions are simpler. However, for $\hat{\Delta}_n$, the scenario is dominated by $\hat{\nu}_n I_p$ where

$$\hat{\nu}_n = n^{-1} \Sigma_{i=1}^{n} \| X_i - \hat{\theta}_n \|^{-1} \tag{3.5}$$

which has unbounded influence of the $\underset{\sim}{X}_i$ "close to" $\hat{\underset{\sim}{\theta}}_n$ and thereby may have highly inflationary effect (i.e., upward bias) on the estimator $\hat{\nu}_n$. For this reason the classical "delta-method" of estimating $\underset{\sim}{\Gamma}_F$, as has been posed in (3.1) through (3.3), may not work out well, particularly, for small values of p. To make this point clear, we consider the case where F is a spherically symmetric d.f., so that by some routine analysis, we have

$$\underset{\sim}{\Delta}_F = \frac{p-1}{p} \nu_F \underset{\sim}{I}_p , \quad \underset{\sim}{\Sigma}_F = p^{-1} \underset{\sim}{I}_p \quad \text{and} \quad \underset{\sim}{\Gamma}_F = p(p-1)^{-2} \nu_F^{-2} \underset{\sim}{I}_p . \tag{3.6}$$

Thus, the problem reduces to that of estimating ν_F^{-2}. Recall that in (3.5), values of $\underset{\sim}{X}_i$ "close to" $\hat{\underset{\sim}{\theta}}_n$ will have inflationary effect on $\hat{\nu}_n$, so that $\hat{\nu}_n^{-2}$ may have serious downward bias on an estimator of ν_F^{-2}. Let us consider an ϵ_n-neighborhood of $\hat{\underset{\sim}{\theta}}_n$:

$$J_n = \{\underset{\sim}{X}_i: \ \| \underset{\sim}{X}_i - \hat{\underset{\sim}{\theta}}_n \| \le \epsilon_n , \ 1 \le i \le n\} , \tag{3.7}$$

where ϵ_n is positive and it converges to 0 at a suitable rate as $n \to \infty$. Let $\hat{\pi}_n$ be the sample measure of J_n (i.e., n^{-1} cardinality of the set J_n). Then the contribution of the r.v.'s in J_n towards $\hat{\nu}_n$ is bounded from below by $\hat{\pi}_n \epsilon_n^{-1}$, and depending on the choice of $\{\epsilon_n\}$ this may be made nonvanishing (stochastically) in the limit. Therefore the usual delta method of estimating ν_F^{-2} (or $\underset{\sim}{\Gamma}_F$) may result in considerable bias particularly for $p = 2$ or 3. As such, we examine the performance of the two other resampling methods in estimating $\underset{\sim}{\Gamma}_F$.

4. JACKKNIFING VS. BOOTSTRAPPING

The classical jackknife does not work out in the univariate case, although delete-d jackknifing, for d not too small, works out well. In the multivariate case, the differentiability picture is different, and hence, it is quite natural to inquire: How far jackknife can be prescribed in the estimation of $\underset{\sim}{\Gamma}_F$?

Let $\hat{\underset{\sim}{\theta}}_n$ be defined as in (1.3), and based on $\underset{\sim}{X}_1, \cdots, \underset{\sim}{X}_{i-1}, \underset{\sim}{X}_{i+1}, \cdots, \underset{\sim}{X}_n$, let $\hat{\underset{\sim}{\theta}}_{n-1}^{(i)}$ be the corresponding L_1-norm estimator, for $i = 1, \ldots, n$. Then let the pseudovalues be defined by

$$\hat{\underset{\sim}{\theta}}_{n,\,i} = n\hat{\underset{\sim}{\theta}}_n - (n-1)\, \hat{\underset{\sim}{\theta}}_{n-1}^{(i)} , \quad i = 1, \ldots, n, \tag{4.1}$$

and let

$$\hat{\underset{\sim}{\theta}}_{nJ} = n^{-1} \Sigma_{i=1}^{n} \underset{\sim}{\theta}_{n,\,i} , \tag{4.2}$$

$$\underset{\sim}{V}_{nJ} = (n-1)^{-1} \Sigma_{i=1}^{n} (\hat{\underset{\sim}{\theta}}_{n,\,i} - \hat{\underset{\sim}{\theta}}_{nJ})\, (\hat{\underset{\sim}{\theta}}_{n,\,i} - \hat{\underset{\sim}{\theta}}_{nJ})' \tag{4.3}$$

be respectively the jackknifed version of $\hat{\underset{\sim}{\theta}}_n$ and the jackknifed variance matrix estimator. Further note that

$$\hat{\underset{\sim}{\theta}}_{nJ} = \hat{\underset{\sim}{\theta}}_n + (n-1)\{n^{-1} \Sigma_{i=1}^{n} (\hat{\underset{\sim}{\theta}}_n - \hat{\underset{\sim}{\theta}}_{n-1}^{(i)})\} , \tag{4.4}$$

where in a regular case (when the bias of $\hat{\theta}_n$ is $O(n^{-1})$), the second term on the right hand side of (4.4) adjusts it further, and reduces the bias of $\hat{\theta}_{nJ}$ to $o(n^{-1})$. In the current context, particularly, for small values of p, the bias of $\hat{\theta}_n$ may not be $O(n^{-1})$, and hence, the bias reduction role of jackknifing may not be that effective. Therefore, we mainly concentrate on the variance estimation role of jackknifing i.e., on the convergence properties of V_{nJ}.

We define $M_n(t)$ as in (2.1), and note that for every i: $1 \le i \le n$,

$$M_n(\hat{\theta}^{(i)}_{n-1} - \theta) = n^{-1/2} \| X_i - \hat{\theta}^{(i)}_{n-1} \|^{-1} (X_i - \hat{\theta}^{(i)}_{n-1}) , \tag{4.5}$$

so that

$$\| M_n(\hat{\theta}^{(i)}_{n-1} - \theta) \| = n^{-1/2}, \quad \forall\, i = 1, \ldots, n, \tag{4.6}$$

and by (1.4), $M_n(\hat{\theta}_n - \theta) = 0$. Thus, proceeding as in the proof of Theorem 2.1, we may verify that as $n \to \infty$

$$\max_{i \le i \le n} \| \hat{\theta}^{(i)}_{n-1} - \hat{\theta}_n \| \to 0 \text{ a.s.}; \quad \| \hat{\theta}_n - \theta \| \to 0 \text{ a.s.} \tag{4.7}$$

Letting $\hat{\theta}_n = \theta + t$ in (3.2), we define the corresponding matrix by $\hat{\Delta}_n(t)$ and for every k(= 1, ..., n), $t \epsilon R^p$, and let

$$V^{(k)}_{n-1}(t) = (n-1)^{-1} \Sigma^n_{i=1} \| X_i - \theta - t \|^{-1} \{ I - \| X_i - \theta - t \|^{-2} (X_i - \theta - t)(X_i - \theta - t)' \} . \tag{4.8}$$

Then, note that for every $n \ge p$,

$$E\{\hat{\Delta}_n(t) \mid \mathfrak{C}_{n+1}\} = \hat{\Delta}_{n+1}(t) , \quad \text{a.e.,} \ \forall\, t \in T \subset R^p , \tag{4.9}$$

(where $\mathfrak{C}_n$ is the tail sigma field), so that for every T,

$$\{\sup \{ \| \hat{\Delta}_n(t) - \hat{\Delta}_n(0) \| : t \in T\} , \mathfrak{C}_n ; n \ge p\} \text{ is a nonnegative reversed sub-martingale.} \tag{4.10}$$

Allowing T to be a small neighborhood of 0 and making use of (4.7), (4.10) and (3.4), it follows that as $n \to \infty$

$$\max_{1 \le k \le n} \{ \sup_{t \in T} \| V^{(k)}_{n-1}(t) - \hat{\Delta}_n(t) \| \} \xrightarrow{P} 0 , \tag{4.11}$$

$$\sup_{t \in T} \| \hat{\Delta}_n(t) - \hat{\Delta}_n(0) \| \xrightarrow{P} 0 , \tag{4.12}$$

when $\text{diam}(T) \to 0$. Thus, if we make a first order Taylor's expansion of $M_n(\hat{\theta}^{(k)}_{n-1} - \theta)$ around $\hat{\theta}_n - \theta$, and make use of (4.7), (4.8), (4.11), (4.12) and (4.1), we obtain that as $n \to \infty$,

$$\max_{1 \leq k \leq n} \| \hat{\Delta}_n (\hat{\theta}_{n,k} - \hat{\theta}_n) - \| X_k - \hat{\theta}_n \|^{-1} (X_k - \hat{\theta}_n) \| \xrightarrow{P} 0 , \tag{4.13}$$

so that

$$\| V_{nJ} - \hat{\Delta}_n^{-1} \hat{\Sigma}_n \hat{\Delta}_n^{-1} \| \xrightarrow{P} 0 , \text{ as } n \to \infty . \tag{4.14}$$

Therefore, V_{nJ} converges in probability to Γ_F, as $n \to \infty$. In passing, we may remark that intuitively the reasons that jackknifing works in the current context are the following:

(i) Under the assumed regularity conditions, $X_1, \ldots, X_n$ are all distinct, with probability 1, and

(ii) $P\{\hat{\theta}_n = X_k \text{ for some k: } 1 \leq k \leq n\} = 0$.

These conditions, in turn, imply that $P\{\hat{\theta}_{n,k} = X_i$ for some $i = 1, \ldots, n$ and $k = 1, \ldots, n\} = 0$, and this controls the inflationary effect to a larger extent. On the other hand, faced with the remark on the order of bias of $\hat{\theta}_n$ made after (4.4), it seems more desirable to use delete-d jackknifing, for some $d \geq 1$.

For every $\underline{i} = (i_1, \ldots, i_d)$ of distinct indices [out of $(1, \ldots, n)$], we define

$$\hat{\theta}^{(d)}_{n, \underline{i}} = d^{-1}\{n\hat{\theta}_n - (n-d)\, \hat{\theta}^{(\underline{i})}_{n-d}\} , \quad \underline{i} \in I , \tag{4.15}$$

where I is the set of all $\binom{n}{d}$ realizations of $\underline{i}$ and $\hat{\theta}^{(\underline{i})}_{n-d}$ is the L_1-norm estimator of θ based on the X_j , $j \in \{1, \ldots, n\} \setminus \underline{i}$. Let then

$$\hat{\theta}^{(d)}_{nJ} = \binom{n}{d}^{-1} \Sigma_{\underline{i} \epsilon I}\, \hat{\theta}^{(d)}_{n, \underline{i}} \qquad \text{(delete-d Jackknifed version)} \tag{4.16}$$

and

$$V^*_{n,\, d} = \binom{n}{d}^{-1} \Sigma_{\underline{i} \epsilon I} (\hat{\theta}^{(d)}_{n, \underline{i}} - \hat{\theta}^{(d)}_{n, J}) (\hat{\theta}^{(d)}_{n, \underline{i}} - \hat{\theta}^{(d)}_{n, J})' . \tag{4.17}$$

Then the delete-d jackknifed variance (matrix) estimator is

$$V^{(d)}_{n, J} = (n-d)^{-1} (n-1)\, d V^*_{n,\, d} . \tag{4.18}$$

By similar manipulations it can be shown that $V^{(d)}_{n, J}$ converges in probability to Γ_F as $n \to \infty$, and for d not too small, but d/n small, it has better robustness properties then $V_{n,\, J}$.

Let us examine the bootstrap method for the estimation of Γ_F . Note that one has to draw a sample of n units with replacement (SRSWR) from $X_1, \ldots, X_n$. If we denote these observations by $X_1^*, \ldots, X_n^*$, respectively, then we solve for (1.4) with the X_i replaced by X_i^*, $i \leq n$. The corresponding estimator is denoted by $\hat{\theta}_n^*$. Suppose now that we draw a large number (say, M) of such (conditionally) independent samples of size n each from $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$, and denote the resulting L_1-norm estimators by $\hat{\theta}^*_{n,1}, \ldots, \hat{\theta}^*_{n,M}$, respectively. Let then

$$V^*_{n,B} = nM^{-1} \Sigma_{i=1}^{M} (\hat{\theta}_{n,i} - \hat{\theta}_n)(\hat{\theta}^*_{n,i} - \hat{\theta}_n)' \tag{4.19}$$

be the bootstrap estimator of Γ_F . We are naturally tempted to follow the same type of manipulations as in (4.5) through (4.14) with the X_i and θ replaced by X^*_i and $\hat{\theta}_n$, respectively. In this context, parallel to (4.7), we may need to show that

$$\max_{1 \le k \le M} \| \hat{\theta}^*_{n,k} - \hat{\theta}_n \| \to 0 \text{ , in probability, as } n \to \infty \text{ .} \tag{4.20}$$

Since the $\hat{\theta}^*_{n,k}$ are conditionally (given $X_1, \ldots, X_n$) i.i.d.r.v.'s, and usually M is taken as O(n), a set of sufficient conditions for (4.20) to hold is that

$$n\, E\{ \| \hat{\theta}^*_n - \hat{\theta}_n \|^2 \mid X_1, \ldots, X_n \} = O(1) \text{ a.e.;} \tag{4.21}$$

$$n^{-1} \Sigma_{i=1}^{n} \| X^*_i - \hat{\theta}_n \|^{-1} - \hat{\nu}_n \xrightarrow{P} 0 \text{ , as } n \to \infty \text{ ,} \tag{4.22}$$

where $\hat{\nu}_n$ is defined by (3.5). Actually, (4.22) stems from a conditional version of (3.4) [given $X_1, \ldots, X_n$] wherein θ and X are replaced by $\hat{\theta}_n$ and X^*, respectively. But $\hat{\nu}^*_n = n^{-1} \Sigma_{i=1}^{n} \| X^*_i - \hat{\theta}^*_n \|^{-1}$ is very vulnerable to the resampling process inherent in the classical bootstrapping. To make this point clear, note that for every k: $1 \le k \le n$,

$$P\{X^*_1 = \cdots = X^*_n = X_k \mid X_1, \ldots, X_n\} = n^{-n}, \tag{4.23}$$

and for $X^*_1 = \cdots = X^*_n = X_k$, $\hat{\theta}^*_n = X_k$, so that $\hat{\nu}^*_n = +\infty$. Thus, given $X_1, \ldots, X_n$, with a conditional probability $> n^{-(n-1)}$, $\hat{\nu}^*_n = +\infty$. In the other extreme case where $X^*_1, \ldots, X^*_n$ are all distinct (i.e., a permutation of $X_1, \cdots, X_n$), $\hat{\theta}^*_n = \hat{\theta}_n$ and $\hat{\nu}^*_n = \hat{\nu}_n$; but this has a probability $(n!)/n^n = n^{-(n-1)}(n-1)! \to 0$, as $n \to \infty$. Also, over the set of realizations of $X^*_1, \ldots, X^*_n$ for which $\hat{\theta}^*_n$ coincides with some X^*_r (which, in turn, is some X_k , $1 \le k \le n$), $\hat{\nu}^*_n = +\infty$. Even if $\hat{\theta}^*_n$ is not exactly equal to some X^*_r , but is arbitrarily close to some X^*_r , $\hat{\nu}^*_r$ can be arbitrarily large, and this possibility can not be ruled out, even for large n, as the conditional distribution of $X^*_1, \ldots, X^*_n$, given $X_1, \ldots, X_n$, is generated by the n^n (conditionally) equally likely realizations $\{X_{i_1}, \ldots, X_{i_n}\}$, where each i_j can take on the values 1, 2, …, n with the common conditional probability n^{-1}. Thus, the vulnerability of the bootstrap method in this L_1-norm estimation problem is mainly due to the discreteness of the conditional distribution of X^*, given $X_1, \ldots, X_n$, which allows $\hat{\nu}^*_n$ to be $+\infty$ with a positive probability (however small it may be for large n), and this, in turn, makes $V^*_{n,B}$ in (4.19) generally underestimating Γ_F. This type of bias of bootstrap variance estimators is quite common for the univariate case too; see for example, Huang [2] for sample median in the discrete case, and Stangenhaus [8] for L_1-regression. Stangenhaus [8] proposed a smoothed bootstrap to overcome the problem of underestimating the asymptotic variance. Her procedure may be extended to the multivariate case to make similar adjustments. Other modifications of the classical bootstrap method may also work out better, and these are under active study.

REFERENCES

[1] Bai, Z. D., Chen, X. R., Miao, B. Q. and Rao, C. R. (1990). Asymptotic theory of least distance estimates in multivariate analysis. Statistics 21, 503-519.

[2] Huang, J. S. (1991). Estimating the variance of the sample median, discrete case. Statistics & Probability Letters 11, 291-298.

[3] Jurečková, J. and Sen, P. K. (1987). A second order asymptotic distributional representation of M-estimators with discontinuous score functions. Annals of Probability 15, 814-823.

[4] Jurečková, J. and Sen, P. K. (1990). Effect of the initial estimator on the asymptotic behavior of one-step M-estimators. Annals of the Institute of Statistical Mathematics 42, 345-347.

[5] Sen, P. K. (1988a). Functional jackknifing: rationality and general asymptotics. Annals of Statistics 16, 450-469.

[6] Sen, P. K. (1988b). Functional approaches in resampling plans: A review of some recent developments. Sankhya, Series A 50, 394-435.

[7] Singer, J. M. and Sen, P. K. (1985). M-methods in multivariate linear models. Journal of Multivariate Analysis 17, 168-184.

[8] Stangenhaus, G. (1987). Bootstrap and inference procedure for L_1-regression. In Statistical Data Analysis Based on the L_1-norm and Related Methods, Y. Dodge (editor), North Holland, pp. 323-332.

Departments of Biostatistics and Statistics
University of North Carolina
Chapel Hill, NC 27599-7400, USA