RESAMPLING METHODS FOR THE EXTREMA OF CERTAIN SAMPLE FUNCTIONS

by

Pranab Kumar Sen

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

Institute of Statistics Mimeo Series No. 1896

March 1992

RESAMPLING METHODS FOR THE EXTREMA OF CERTAIN SAMPLE FUNCTIONS

PRANAB KUMAR SEN

University of North Carolina at Chapel Hill

For a general class of extrema of sample functions, including bundle strength of filaments, for variance estimation, the classical delta method along with jackknifing and bootstrapping are critically examined and their relative merits and demerits are discussed.

1. INTRODUCTION

We motivate our study with the following simple model. Consider a bundle of n parallel filaments whose individual breaking strengths are denoted by $X_1, \ldots, X_n$. These X_i are assumed to be independent and identically distributed (i.i.d.) nonnegative random variables (r.v.) having a continuous distribution function (d.f.) F, defined on $R^+ = [0,\infty)$. Let $X_{n:1} < \ldots < X_{n:n}$ be the associated order statistics, and as in Daniels (1945), define

$$D_n = \max\{ (n-i+1)X_{n:i} \; ; \; 1 \leq i \leq n \} \tag{1.1}$$

as the *bundle strength of filaments*. Daniels prescribed a very elaborate analysis leading to the asymptotic normality of a normalized version of D_n. Based on the behavior of the empirical d.f. (e.d.f.) $F_n(x) = n^{-1}\Sigma_{i=1}^n I(X_i \leq x)$, $x \in R$,a greatly simplified approach to this study was initiated by Sen, Bhattacharyya and Suh (1973). We may note in this context that

$$\begin{aligned} Z_n = n^{-1}D_n &= \max\{ [1 - (i-1)/n]X_{n:i} : 1 \leq i \leq n \} \\ &= \sup\{ [1-F_n(x)]x : x \in R^+ \} , \; n \geq 1, \end{aligned} \tag{1.2}$$

so that the asymptotic behavior of F_n may provide us with all the necessary tools for the study of the same for Z_n. Keeping this in mind and proceeding as in Bhattacharyya and Sen (1976) and Sen (1976, 1981), we may conceive of a sequence $\{X_i ; i \geq 1\}$ of i.i.d.r.v.'s with a continuous d.f. F, defined on R^p, for some $p \geq 1$, along with a smooth function $\psi : R^p \times [0,1] \to R$, and consider a functional

$$\theta = \theta(F) = \sup\{ \psi(x,F(x)) : x \in A \subset R^p \}. \tag{1.3}$$

AMS Subject Classifications Nos : 62E20, 62E99

Key Words & Phrases : Asymptotic normality; asymptotic variance; bias; bootstrapping; bundle strength of filaments; delta methos; jackknifing.

Let F_n be the e.d.f. based on $X_1,\dots,X_n$, and define

$$\hat{\theta}_n = \sup\{\ \psi(x,F_n(x)) : x \in A\},\quad n \geq n_0 . \tag{1.4}$$

We assume that there exists a unique x_0, an interior point of A, such that

$$-\infty < \theta = \psi(x_0,F(x_0)) < +\infty \quad \text{and } 0 < \pi_0 = F(x_0) < 1. \tag{1.5}$$

Also, we assume that for for every $\varepsilon > 0$, there exists an $\eta > 0$, such that

$$\psi(x,F(x)) < \theta - \varepsilon \quad , \text{for every } x : ||\ x - x_0|| > \eta . \tag{1.6}$$

Moreover, we assume that

$$\sup_{x \in A} |\ \psi(x,F_n(x)) - \psi(x,F(x))| \to 0, \text{ on the set } ||F_n-F|| \to 0. \tag{1.7}$$

In many cases of practical interest, (1.6) may be verified by incorporating Hadamard or compact continuity of functionals [viz., Sen (1988)]. Finally,we assume that $\psi_{01}(x,y) = (\partial/\partial y)\psi(x,y)$ exists for every $x \in A$ and $y \in (0,1)$, and further, there exists a neighborhood of (x_0,π_0) in which $\psi_{01}(.)$ is uniformly continuous; we denote by

$$\xi = \psi_{01}(x_0,\pi_0) \quad \text{and} \quad \gamma^2 = \xi^2\pi_0(\ 1 - \pi_0). \tag{1.8}$$

First, we may note that by (1.3) and (1.4), for every $n \geq n_0$,

$$|\ \hat{\theta}_n - \theta\ | \leq \sup_{x \in A} |\ \psi(x,F_n(x)) - \psi(x,F(x))|\ , \tag{1.9}$$

so that by (1.7) and the fact (that by the Glivenko-Cantelli lemma) $||F_n-F|| \to 0$ almost surely (a.s.) as $n\to \infty$, we obtain that

$$|\ \hat{\theta}_n - \theta\ | \to 0 \text{ a.s., as } n \to \infty . \tag{1.10}$$

Also, if we set

$$\hat{\theta}_n = \psi(\hat{x}_{on}, \hat{\pi}_{on}) \quad \text{where} \quad \hat{\pi}_{on} = F_n(\hat{x}_{on})\ , \tag{1.11}$$

then, by (1.10), (1.11) and the a.s. convergence of $||F_n - F||$ to 0 , we have

$$||x_{on} - x_0\ || \to 0 \text{ a.s., and } \hat{\pi}_{on} - \pi_0 \to 0 \text{ a.s., as } n \to \infty. \tag{1.12}$$

Further, by (1.3) and (1.4), $\psi(x_0,F_n(x_0)) - \psi(x_0,F(x_0)) \leq \hat{\theta}_n - \theta \leq \psi(\hat{x}_{on},\hat{\pi}_{on}) - \psi(\hat{x}_{on},F(\hat{x}_{on}))$, with probability 1. Moreover, if we incorporate the compactness part of the weak convergence of $n^{1/2}(F_n - F)$, we may claim that for every $\varepsilon > 0$, there exist an $\eta > 0$ and a sample size n_0 (= $n_0(\varepsilon,\eta)$), such that

$$P\{\sup_{||x-x_0|| <\varepsilon} n^{1/2}|F_n(x)-F(x)-F_n(x_0)+F(x_0)| > \varepsilon\ \} < \eta \quad , \forall\ n \geq n_0. \tag{1.13}$$

At this stage, we make use of (1.12) and the uniform continuity of $\psi_{01}(.)$ in a neighborhood of (x_0,π_0), and obtain from the above analysis that as n increases,

$$n^{1/2}(\ \hat{\theta}_n - \theta\ ) \xrightarrow{\mathcal{D}} N(\ 0, \gamma^2\ ). \tag{1.14}$$

This result extends Theorem 8.1.6 of Sen(1981) under slightly less stringent regularity conditions. In order to draw statistical conclusions on θ based on $\hat{\theta}_n$, the use of (1.14) rests on the estimation of γ^2. Moreover, $\hat{\theta}_n$ is a highly non-linear estimator, and hence, its bias may not be of the order n^{-1}. As such, there may be a need to study the nature of bias , so as to make plausible corrections for moderate sample sizes. We shall look into both these variance estimation and bias estimation problems.

2. ORDER OF ASYMPTOTIC BIAS

We have already noticed that with probability one,

$$\psi(x_o,F_n(x_o)) - \psi(x_o,F(x_o)) \leq \hat{\theta}_n - \theta \leq \psi(\hat{x}_{on},F_n(\hat{x}_{on})) - \psi(\hat{x}_{on},F(\hat{x}_{on})). \quad (2.1)$$

In the particular case of the bundle strength of filaments, $\psi(x,y) = x(1-y)$, $x \in R^+$ and $y \in [0,1]$, so that the left hand side of (2.1) is equal to $x_o[F(x_o) - F_n(x_o)]$ which has mean zero. Hence, $\hat{\theta}_n$ has a nonnegative bias. In fact, in this special case, by the use of the reverse martingale property of $\{F_n\}$, it has been shown in Sen et al. (1973) that $\{\hat{\theta}_n\}$ is a reverse submartingale, and further, the bias of $\hat{\theta}_n$ is $o(n^{-1/2})$, as $n \to \infty$. Even in this special case, direct manipulation of the right hand side of (2.1) seemed to be quite complicated (as $\hat{x}_{on}$ is stochastic), and a bias of the order n^{-1} did not appear to be evident !

To study the bias term in a general setup, we rewrite (2.1) as

$$\begin{aligned} 0 &\leq \{ \hat{\theta}_n - \theta - [\psi(x_o,F_n(x_o)) - \psi(x_o,F(x_o))] \} \qquad (2.2) \\ &\leq \{ [\psi(\hat{x}_{on},F_n(\hat{x}_{on})) - \psi(\hat{x}_{on},F(\hat{x}_{on}))] - [\psi(x_o,F_n(x_o)) - \psi(x_o,F(x_o))]\} . \end{aligned}$$

Given enough smoothness conditions on $\psi(..)$, we may be tempted to make use of a Bahadur (1966) type representation for the right hand side of (2.2) wherein we may need to use suitable rates of convergence in (1.12) and (1.13). Further, as we need to take expectations , we may have to use the Duttweiler (1973) approach to the Bahadur representation along with Theorem 7.3.1 of Sen (1981) to cope with a slower rate than $n^{-1/2}$ (for $\hat{x}_{on} \to x_o$). The end product is that typically the bias of ($\hat{\theta}_n - \theta$) is $O(n^{-\lambda})$, for some $\lambda : 1/2 < \lambda \leq 3/4$. In this context, we may note that the Hadamard continuity in (1.7) does not necessarily insure the Hadamard differentiability, and hence, the usual (first or second order) differentiability properties of the functional under review may not hold. Expecting a second order differentiability property to hold would restrict the class of $\psi(..)$ and hamper the generality of the result in (1.14). Also, in the negation of such differentiability, the classical jackknifing may not work out well in reducing the bias to a lower order. For this reason, we need to explore other possibilities. Among these the sub-sampling method considered by Sen (1990) may work out well.

In this context, we may note that if for a statistic T_n based on n observations,

$$E(T_n) = \theta + a(F)n^{-\lambda} + o(n^{-\lambda}), \text{ for some } \lambda : 0 < \lambda \leq 1, \tag{2.3}$$

where $a(F)$ is an unknown functional of the df F, then

$$E\{ nT_n - (n-1)T_{n-1}\} = \theta + (1-\lambda)n^{-\lambda}.a(F) + o(n^{-\lambda}), \tag{2.4}$$

so that for $\lambda < 1$, the order of the bias term remains the same, although its contribution is discounted by the factor $(1-\lambda)$. This very simple observation implies that for an order of bias $O(n^{-\lambda})$, $\lambda < 1$, there may not be enough incentive in using the classical jackknife for bias reduction, although its utility in estimating the asymptotic variance may still remain in tact. The situation is no better for the classical bootstrap. Since the bootstrap sampling allows a possible duplication of some of the units in the original sample (with a positive probability) , the effect of ties arising in this manner may actually make the bias term even worse ! For this reason, we may find it convenoent to use a delete-d jackknifing method (where d is not too small) as a better compromise in controlling the bias term without making any real difference in the asymptotic variance term. We may refer to some of the details in Sen (1989).

3. RESAMPLING PLANS FOR VARIANCE ESTIMATION

First, we consider a naive estimator of γ^2. Using (1.8), (1.10), (1.11) and (1.12), we propose

$$\hat{\gamma}_n^2 = \hat{\xi}_n^2 \hat{\pi}_{on}(1 - \hat{\pi}_{on}) \tag{3.1}$$

as an estimator of γ^2 , where we let

$$\hat{\xi}_n = \psi_{01}(\hat{x}_{on} , \hat{\pi}_{on}). \tag{3.2}$$

This is actually based on the classical delta-method, and the estimator is consistent under the assumed regularity conditions. However, as we have noticed that the original estimator $\hat{\theta}_n$ in (1.4) is highly non-linear, the estimator $\hat{\gamma}_n$ in (3.1) may inherit the sensitivity of the delta-method to basic nonlinearity of the functional,and as such may have significant bias component. For this reason, we may like to explore alternative methods based on suitable resampling plans with due emphasis on the classical jackknifing and bootstrapping methods.

Let $\underset{\sim}{X}_{n-1}^{(i)}$ be the sample of size n-1 obtained by deleting the ith observation (X_i) from the base sample of size n, for i=1,...,n. We denote the empirical d.f. associated with $\underset{\sim}{X}_{n-1}^{(i)}$ by $F_{n-1}^{(i)}$,and the estimator of θ based on this subsample is denoted by $\hat{\theta}_{n-1}^{(i)}$, for i= 1,...,n. In the classical jackknife method, we then introduce the pseudo-values $\hat{\theta}_{n,i}$, i= 1,..., n, as

$$\hat{\theta}_{n,i} = n\,\hat{\theta}_n - (n-1)\,\hat{\theta}_{n-1}^{(i)} \quad , \quad \text{for } i = 1,\ldots,n. \tag{3.3}$$

Then the classical jackknifed version of the estimator $\hat{\theta}_n$ is given by

$$\begin{aligned}\hat{\theta}_{nJ} &= n^{-1}\,\Sigma_{i=1}^{n}\ \hat{\theta}_{n,i} \\ &= \hat{\theta}_n + (n-1)\{\ n^{-1}\,\Sigma_{i=1}^{n}\ (\ \hat{\theta}_{n-1}^{(i)} - \hat{\theta}_n\ )\}.\end{aligned} \tag{3.4}$$

It follows from (2.3) and (2.4) that the bias of the jackknifed version is $O(n^{-\lambda})$ where λ is greater than 1/2, but not more than 3/4. In this sense, jackknifing has reduced the bias to some extent, but not to any lower order of magnitude. Let us also define

$$\begin{aligned}V_{nJ} &= (n-1)^{-1}\,\Sigma_{i=1}^{n}\{\ \hat{\theta}_{n,i} - \hat{\theta}_{nJ}\ \}^2 \\ &= (n-1)\,\Sigma_{i=1}^{n}(\ \hat{\theta}_{n-1}^{(i)} - n^{-1}\,\Sigma_{j=1}^{n}\ \hat{\theta}_{n-1}^{(j)}\ )^2\end{aligned} \tag{3.5}$$

as the jackknifed variance estimator.Since the subsample (nonlinear) estimators $\hat{\theta}_{n-1}^{(i)}$, i=1,...,n and $\hat{\theta}_n$ are directly involved in (3.5), V_{nJ} is expected to be less sensitive to the inherent nonlinearity of the functional than the classical delta method. Let $C_n = C(X_{n:1},\ldots,X_{n:n};X_{n+j},j \geq 1\ )$, $n \geq 1$ be the sequence of tail sigma field. Then, we may rewrite (3.5) as

$$V_{nJ} = n(n-1)E[\{\ \hat{\theta}_{n-1} - E(\ \hat{\theta}_{n-1} |\ C_n\ )\ \}^2 \mid C_n\ ] \tag{3.6}$$

so that a sufficient condition for the consistency of V_{nJ} is the a.s. convergence of the standardized conditional variance of $\hat{\theta}_n$, given C_{n+1} , as $n \to \infty$.

In the particular case of the bundle strength of filaments, we have already noticed that $\{\hat{\theta}_n;n \geq 1\}$ is a reversed submartingale, and hence, we may even link (3.6) as a vital component towards the asymptotic normality result in (1.14). Motivated by this, our goal may be to approximate the sequence $\{\hat{\theta}_n;\ n \geq 1\}$ by a reversed martingale, and use this characterization in the convergence of V_{nJ} in (3.5) or (3.6). For the subsample $\underset{\sim}{X}_{n-1}^{(i)}$, we define $\hat{\theta}_{n-1}^{(i)}$, $\hat{x}_{o,n-1}^{(i)}$ and $\hat{\pi}_{o,n-1}^{(i)}$ as in (1.11), for every i(=1,...,n). Then, note that

$$\max\{\ \sup[|\ F_{n-1}^{(i)}(x) - F_n(x)|\ :\ x \in R]\ :\ 1 \leq i \leq n\} = n^{-1}\ , \tag{3.7}$$

with probability one, and hence, we can easily extend (1.12) as follows: As $n \to \infty$,

$$\max_{1\leq i\leq n} ||\ x_{o,n-1}^{(i)} - x_o\ || \to 0 \text{ a.s., and } \max_{1\leq i\leq n} |\ \hat{\pi}_{o,n-1}^{(i)} - \pi_o| \to 0 \text{ a.s.} \tag{3.8}$$

This enables us to make use of the uniform continuity result on $\psi_{01}(.)$ for every subsample $\underset{\sim}{X}_{n-1}^{(i)}$, i=1,...,n, when n is large. Note that for every i(=1,...,n),

$$\hat{\theta}_{n-1}^{(i)} = \psi(\ \hat{x}_{o,n-1}^{(i)},\ \ F_{n-1}^{(i)}(\hat{x}_{o,n-1}^{(i)}))\ , \tag{3.9}$$

so that

$$\psi(\hat{x}_{on}, F_{n-1}^{(i)}(\hat{x}_{on})) - \psi(\hat{x}_{on}, F_n(\hat{x}_{on})) \leq \hat{\theta}_{n-1}^{(i)} - \hat{\theta}_n \qquad (3.10)$$
$$\leq \psi(\hat{x}_{o,n-1}^{(i)}, F_{n-1}^{(i)}(\hat{x}_{o,n-1}^{(i)})) - \psi(\hat{x}_{o,n-1}^{(i)}, F_n(\hat{x}_{0,n-1}^{(i)})) .$$

On the other hand, for every $x \in R^p$,

$$F_{n-1}^{(i)}(x) - F_n(x) = (n-1)^{-1}\{ nF_n(x) - I(X_i \leq x) \} - F_n(x)$$
$$= (n-1)^{-1}\{ F_n(x) - I(X_i \leq x)\} , i = 1,\dots,n. \qquad (3.11)$$

Thus, using the uniform continuity of $\psi_{01}(.)$ in a neighborhood of (x_o, π_o) along with (3.8), we conclude that the left hand side of (3.10) behaves as

$$\psi_{01}(\hat{x}_{on}, F_n(\hat{x}_{on}))\{ (n-1)^{-1}[F_n(\hat{x}_{on}) - I(X_i \leq \hat{x}_{on})]\} + o_p(n^{-1}), \qquad (3.12)$$

uniformly in $i = 1,\dots,n$. In a similar manner, the right hand side of (3.10) can be expressed as

$$\psi_{01}(\hat{x}_{o,n-1}^{(i)}, F_n(\hat{x}_{o,n-1}^{(i)}))\{ (n-1)^{-1}[F_n(\hat{x}_{o,n-1}^{(i)}) - I(X_i \leq \hat{x}_{o,n-1}^{(i)})]\} + o_p(n^{-1}), \qquad (3.13)$$

for every i $(=1,\dots,n)$. Incorporating (3.8) along with the uniform continuity of $\psi_{01}(.)$ in a neighborhood of (x_o,π_o), the first factor in (3.12) as well as (3.13) can be replaced by $\psi_{01}(x_o,\pi_o)$, and the difference can be absorbed in the second term i.e., with $o_p(n^{-1})$, uniformly in i $(=1,\dots,n)$. Secondly, if we sum over i $(=1,\dots,n)$, the second factor of the first term in (3.12) becomes identically equal to 0, so that, by (3.10) and (3.12), we obtain that

$$n^{-1} \Sigma_{i=1}^n (\hat{\theta}_{n-1}^{(i)} - \hat{\theta}_n) \geq o_p(n^{-1}). \qquad (3.14)$$

Similarly, if we let

$$\hat{x}_{o,n}^{*} : F_n(\hat{x}_{o,n}^{*}) = \max_{1\leq i\leq n} F_n(\hat{x}_{o,n-1}^{(i)}) \text{ and } \hat{x}_{o,n}^{**}: F_n(\hat{x}_{o,n}^{**}) = \min_{1\leq i\leq n} F_n(\hat{x}_{o,n-1}^{(i)}), \qquad (3.15)$$

then

$$F_n(\hat{x}_{o,n}^{**}) - F_n(\hat{x}_{o,n}^{*}) \leq n^{-1}\Sigma_{i=1}^n[F_n(\hat{x}_{o,n-1}^{(i)}) - I(X_i \leq \hat{x}_{o,n-1}^{(i)})]$$
$$\leq F_n(\hat{x}_{o,n}^{*}) - F_n(\hat{x}_{o,n}^{**}), \qquad (3.16)$$

where, by (3.8), both the left and right hand sides of (3.16) are $o(1)$ a.s., so that

$$n^{-1}\Sigma_{i=1}^n (\hat{\theta}_{n-1}^{(i)} - \hat{\theta}_n) \leq o_p(n^{-1}). \qquad (3.17)$$

From (3.14) and (3.17), we obtain that as $n \to \infty$,

$$n^{-1}|\Sigma_{i=1}^n(\hat{\theta}_{n-1}^{(i)} - \hat{\theta}_n)| = o_p(n^{-1}). \qquad (3.18)$$

In fact, in (3.12) and (3.13), the remainder terms are $o(\frac{1}{n})$ a.s., as $n \to \infty$, and hence, in (3.18) too, we may replace $o_p(\frac{1}{n})$ by $o(\frac{1}{n})$ a.s., as $n \to \infty$. As such, by

(3.5), (3.6) and (3.18), we conclude that

$$V_{nJ} = (n-1)^{-1}\Sigma_{j=1}^{n} \left(\hat{\theta}_{n-1}^{(j)} - \hat{\theta}_n \right)^2 + o(1) \text{ a.s., as } n \to \infty. \tag{3.19}$$

This enables us to use again (3.10) through (3.13) with the $o_p(n^{-1})$ terms being replaced by $o(n^{-1})$ a.s., as $n \to \infty$. Towadrs this, using (3.8), we first write (3.13) as

$$\psi_{01}(\hat{x}_{on}, F_n(\hat{x}_{on}))\{ (n-1)^{-1}[F_n(\hat{x}_{o,n-1}^{(i)}) - I(X_i \le \hat{x}_{o,n-1}^{(i)})]\} + o(n^{-1}) \text{ a.s.} \tag{3.20}$$

Secondly, we write

$$F_n(\hat{x}_{o,n-1}^{(i)}) - I(X_i \le \hat{x}_{o,n-1}^{(i)}) = F_n(\hat{x}_{on}) - I(X_i \le \hat{x}_{on}) + \hat{R}_{ni}, \tag{3.21}$$

for every $i(=1,\dots,n)$, where

$$\hat{R}_{ni} = F_n(\hat{x}_{o,n-1}^{(i)}) - F_n(\hat{x}_{on}) - I(X_i \le \hat{x}_{o,n-1}^{(i)}) + I(X_i \le \hat{x}_{on}), \; i=1,\dots,n. \tag{3.22}$$

Note that by (3.15) and (3.22),

$$\begin{aligned} n^{-1}\Sigma_{i=1}^{n} \hat{R}_{ni}^2 &\le 2\, n^{-1}\Sigma_{i=1}^{n}[I(X_i \le \hat{x}_{o,n-1}^{(i)}) - I(X_i \le \hat{x}_{on})]^2 \\ &\quad + 2\, n^{-1}\Sigma_{i=1}^{n}[F_n(\hat{x}_{o,n-1}^{(i)}) - F_n(\hat{x}_{on})]^2 \\ &\le 2\, n^{-1}\Sigma_{i=1}^{n} I(\hat{x}_{o,n}^{**} \le X_i \le \hat{x}_{o,n}^{*}) \\ &\quad + 2[F_n(\hat{x}_{o,n}^{*}) - F_n(\hat{x}_{o,n}^{*})]^2 \\ &= 2[F_n(\hat{x}_{o,n}^{*}) - F_n(\hat{x}_{o,n}^{**})][1 + F_n(\hat{x}_{o,n}^{*}) - F_n(\hat{x}_{o,n}^{**})] \\ &= o(1) \text{ a.s., as } n \to \infty, \end{aligned} \tag{3.23}$$

where in the last step we have made use of (3.8), (3.15) and the Glivenko-Cantelli lemma. Consequently, by (3.12), (3.13), (3.20) - (3.23), we may write V_{nJ} in (3.19) as

$$\begin{aligned} V_{nJ} &= (n-1)^{-1}\Sigma_{i=1}^{n}\psi_{01}^2(\hat{x}_{on},F_n(\hat{x}_{on}))[F_n(\hat{x}_{on}) - I(X_i \le \hat{x}_{on})]^2 + o(1) \text{ a.s.} \\ &= n(n-1)^{-1}\, \psi_{01}^2(\hat{x}_{on},F_n(\hat{x}_{on}))\, F_n(\hat{x}_{on})[1-F_n(\hat{x}_{on})] + o(1) \text{ a.s.} \\ &= n(n-1)^{-1}\, \hat{\gamma}_n^2 + o(1) \text{ a.s., as } n \to \infty, \end{aligned} \tag{3.24}$$

where $\hat{\gamma}_n^2$ is defined by (3.1), and by (1.12) and (3.1), $\hat{\gamma}_n^2$ is a strongly consistent estimator of γ^2, defined by (1.8). Thus, we obtain that

$$V_{nJ} \to \gamma^2 \text{ a.s., as } n \to \infty. \tag{3.25}$$

Although the (strong) consistency of V_{nJ} has been established here through the asymptotic a.s. equivalence result in (3.24) and the a.s. convergence of the naive estimator $\hat{\gamma}_n^2$ to γ^2, as has been noted earlier, V_{nJ} may be much less sensitive to nonlinearity of $\psi(.)$, and hence, should be more robust.

Let us look into the bootstrap variance estimation problem. Corresponding to the given sample $X_1,\dots,X_n$, we define the empirical d.f. F_n as in earlier, and let $X_1^*,\dots,X_n^*$ be n (conditionally) i.i.d.r.v.'s drawn from the d.f. F_n ; we denote the empirical d.f. for the X_i^* by F_n^* . Then, we define

$$\hat{\theta}_n^* = \sup\{ \ \psi(x, F_n^*(x)) : x \in A \subset R^p \ \}. \tag{3.26}$$

For latter use, we write

$$\hat{\theta}_n^* = \psi(\hat{x}_{on}^* , \hat{\pi}_{on}^*) \text{ where } \hat{\pi}_{on}^* = F_n^*(\hat{x}_{on}^*) . \tag{3.27}$$

For some positive integer M (usually large), we draw M such (conditionally independent) bootstrap samples from F_n, and denote the corresponding estimators by

$$\hat{\theta}_{n,1}^* , \dots , \hat{\theta}_{n,M}^* . \tag{3.28}$$

Let then

$$V_{nB}^* = nM^{-1} \Sigma_{i=1}^M (\hat{\theta}_{n,i}^* - \hat{\theta}_n)^2 \tag{3.29}$$

be the bootstrap estimator of the asymptotic variance of $n^{1/2}(\hat{\theta}_n - \theta)$. We like to study the convergence properties of V_{nB}^* and also compare it with the jackknifed variance estimator V_{nJ}.

Recall that by definition,

$$\psi(\hat{x}_{on},F_n^*(\hat{x}_{on})) - \psi(\hat{x}_{on},F_n(\hat{x}_{on})) \le \hat{\theta}_n^* - \hat{\theta}_n \le \psi(\hat{x}_{on}^*,F_n^*(\hat{x}_{on}^*)) - \psi(\hat{x}_{on}^*,F_n(\hat{x}_{on}^*)), \tag{3.30}$$

and further,

$$n^{1/2} || F_n^* - F_n || = O_p(1) \text{ a.e. } (F_n). \tag{3.31}$$

By virtue of (3.31) and the assumed regularity conditions on ψ , it can be shown that as $n \to \infty$,

$$||\hat{x}_{on}^* - \hat{x}_{on} || \to 0 \text{ and } |\hat{\pi}_{on}^* - \hat{\pi}_{on} | \to 0, \text{ in probability,} \tag{3.32}$$

and further, the classical Bahadur (1966) representation of sample quantiles holds for the bootstrap empirical d.f.[see for example, Gangopadhyay and Sen (1990)]. As such, using the uniform continuity of $\psi_{01}(.)$ in a neighborhood of (x_o, π_o), it follows from (3.30) through (3.32) that as $n \to \infty$,

$$n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) - \psi_{01}(\hat{x}_{on},F_n(\hat{x}_{on})) \, n^{1/2}[F_n^*(\hat{x}_{on}) - F_n(\hat{x}_{on})] \to 0, \text{ in probability (a.e. } F_n \text{) ,} \tag{3.33}$$

so that using the conditional i.i.d. nature of the set in (3.28), and the representation in (3.33), we obtain from (3.29) that as $n \to \infty$,

$$V_{nB}^* - \psi_{01}^2(\hat{x}_{on},F_n(\hat{x}_{on}))F_n(\hat{x}_{on})[1-F_n(\hat{x}_{on})] \to 0, \text{ in probability.} \tag{3.34}$$

Therefore, from (3.1), (3.34) and the consistency of the naive estimator $\hat{\gamma}_n^2$, we conclude that under the assumed regularity conditions, as $n \to \infty$,

$$V_{nB}^{*}\ ,\ V_{nJ} \text{ and } \hat{\gamma}_n^2 \text{ are equivalent in probability, and they estimate } \gamma^2 \text{ consistently.} \tag{3.35}$$

It may be remarked here that $\hat{\gamma}_n^2$ and V_{nJ} are both strongly consistent estimators of the asymptotic variance of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta)$, whereas, we have been able to establish only the weak consistency of the bootstrap variance estimator. This is mainly due to the fact that in bootstrapping we consider the conditional law given the base sample, and hence, an a.s. result will not only require the a.s. part on the conditional setup but also on the part of the base sample. Thus, we would have to deal with a double sequence, and hence, we may need extra regularity conditions. On the other hand, for the conditional law, there are some nice recent developments [viz., Gine and Zinn (1990)] which may be incorporated to obtain some uniformity results. However, we shall not probe into these here.

4. PERFORMANCE CHARACTERISTICS UNDER ADDITIONAL SMOOTHNESS CONDITIONS

For simplicity of presentation, we assume that in (1.3)-(1.4), x or the X_i are real valued; the results to follow continue to hold even if $x \in R^p$, for $p \geq 2$. We assume that in a neighborhood of (x_0, π_0), $\psi(x,y)$ has continuous second order partial derivatives (with respect to x,y), and these are bounded. We denote the first order derivatives by $\psi_{10}(.)$ and $\psi_{01}(.)$, while the second order ones are denoted by $\psi_{20}(.)$, $\psi_{11}(.)$ and $\psi_{02}(.)$ respectively. Also, we denote the density function for the d.f. F by f(x), and assume that f(x) is continuous in a neighborhood of x_0. Then, note that by definition of θ in (1.3) and (1.5),

$$\psi_{10}(x_0, \pi_0) + f(x_0)\, \psi_{01}(x_0,\pi_0) = 0 . \tag{4.1}$$

We have already assumed that $\psi_{01}(x_0,\pi_0)$ is nonzero (as otherwise γ^2 will be equal to 0 and we would have a degenerate normal law in (1.14)). Then, by a local expansion of $\psi(x,F(x))$ around $\psi(x_0,F(x_0))$, we obtain that as $x \to x_0$,

$$(x-x_0)\psi_{10}(x_0,F(x_0)) = -[F(x)-F(x_0)]\psi_{01}(x_0,F(x_0)) + O((F(x)-F(x_0))^2). \tag{4.2}$$

On the other hand, by (4.1), and a local expansion of $\psi(x,F(x))$ around $(x_0,F(x_0))$, we obtain that as $x \to x_0$,

$$\psi(x,F(x)) = \psi(x_0,F(x_0)) - O(\ [x - x_0]^2 + [F(x) - F(x_0)]^2\ ). \tag{4.3}$$

Combining (4.2), (4.3) and the Bahadur (1966) representation for $\{ F_n(x) - F(x) - F_n(x_0) + F(x_0) \}$, $|x - x_0| = O(n^{-\beta}(\log n)^{1/2})$, for some $\beta : \beta \leq 1/2$), [see, Theorem 7.3.1 of Sen (1981)] , we obtain that

$$| \hat{x}_{on} - x_o | = O(n^{-1/3}(\log n)^{1/2}) \text{ a.s., as } n \to \infty . \tag{4.4}$$

As such, by an iterated use of Theorem 7.3.1 of Sen (1981), we obtain from (3.1), (4.4) and the assumed boundedness of $\psi_{02}(.)$ in a neighborhood of $(x_o, F(x_o))$ that

$$\hat{\gamma}_n^2 = \gamma^2 + O(n^{-1/3}(\log n)^{1/2}) \quad \text{a.s., as } n \to \infty . \tag{4.5}$$

This exhibits the good performance of the naive estimator $\hat{\gamma}_n$ in (3.1) even when the functional is not so smooth to induce a bias term of the order n^{-1}.

Let us next look into the picture of the jackknifed variance estimator V_{nJ}. As in before (3.7), we introduce the notations $\hat{x}_{o,n-1}^{(i)}$, $F_{n-1}^{(i)}$ etc. for $i=1,\ldots,n$. Using (3.10) along with the fact that $\psi(\hat{x}_{on}, F_n(\hat{x}_{on})) \geq \psi(\hat{x}_{o,n-1}^{(i)}, F_n(\hat{x}_{o,n-1}^{(i)}))$, for every i $(=1,\ldots,n)$, with probability 1, we obtain that for every $i(=1,\ldots,n)$,

$$\begin{aligned} 0 &\leq (\hat{\theta}_{n-1}^{(i)} - \hat{\theta}_n) - [\psi(x_{on}, F_{n-1}^{(i)}(\hat{x}_{on})) - \psi(\hat{x}_{on}, F_n(\hat{x}_{on}))] \\ &\leq \{ \psi(\hat{x}_{o,n-1}^{(i)}, F_{n-1}^{(i)}(\hat{x}_{o,n-1}^{(i)})) - \psi(\hat{x}_{on}, F_{n-1}^{(i)}(\hat{x}_{on})) - \psi(\hat{x}_{o,n-1}^{(i)}, F_n(\hat{x}_{o,n-1}^{(i)})) \\ &\qquad + \psi(\hat{x}_{on}, F_n(\hat{x}_{on})) \}. \end{aligned} \tag{4.6}$$

Making use of the Taylor expansion on the right hand side of (4.6) and the fact that $F_{n-1}^{(i)}(x) - F_n(x) = -\{I(X_i \leq x) - F_n(x)\}/(n-1)$, for every x and i, we obtain that the right hand side of (4.6) can be written as

$$\begin{aligned} &\psi_{01}(\hat{x}_{on}, F_n(\hat{x}_{on}))(n-1)^{-1}\{ I(X_i \leq \hat{x}_{on}) - I(X_i \leq \hat{x}_{o,n-1}^{(i)}) - F_n(\hat{x}_{on}) + \\ &\qquad F_n(\hat{x}_{o,n-1}^{(i)}) \} + o(n^{-1}) \text{ a.s., as } n \to \infty, \end{aligned} \tag{4.7}$$

so that

$$\begin{aligned} &(n-1)[(\hat{\theta}_{n-1}^{(i)} - \hat{\theta}_n) - \{\psi(\hat{x}_{on}, F_{n-1}^{(i)}(\hat{x}_{on})) - \psi(\hat{x}_{on}, F_n(\hat{x}_{on})) \}]^2 \\ &\qquad = \psi_{01}^2 (\hat{x}_{on}, F_n(\hat{x}_{on})) \cdot A_{ni}^2, \end{aligned} \tag{4.8}$$

where

$$\Sigma_{i=1}^n A_{ni}^2 = O(n^{-1/3} \log n) \text{ a.s., as } n \to \infty . \tag{4.9}$$

In the derivation of (4.9), we have made use of (4.1) through (4.4) along with (3.15) and (3.16), whereby we are able to show that $F_n(\hat{x}_{o,n}^{*}) - F_n(\hat{x}_{o,n}^{**}) = O(n^{-1/3} \cdot \log n)$ a.s., as $n \to \infty$. By virtue of (4.8) and (4.9), we obtain that as $n \to \infty$,

$$\begin{aligned} V_{nJ} &= (n-1) \Sigma_{i=1}^n \{ \psi(\hat{x}_{on}, F_{n-1}^{(i)}(\hat{x}_{on})) - \psi(\hat{x}_{on}, F_n(\hat{x}_{on})) \}^2 + O(n^{-1/3} \log n) \text{ a.s.} \\ &= V_{nJ}^* + O(n^{-1/3} \log n) \text{ a.s.,} \end{aligned} \tag{4.10}$$

where by a simple Taylor expansion, it follows that as $n \to \infty$,

$$V_{nJ}^* = \hat{\gamma}_n^2 + O(n^{-1/2} \log n) \text{ a.s..} \tag{4.11}$$

Thus, (4.5) extends to V_{nJ} as well. On the other hand, in (4.10), the nonlinearity of $\psi(.)$ is retained to a greater extent (than in (3.1)), so that V_{nJ}^* and hence

V_{nJ} are likely to be more robust than $\hat{\gamma}_n^2$ when n is not so large.

Let us consider next the bootstrap variance estimator V_{nB}^* defined by (3.29). Note that given F_n, the estimators in (3.28) are conditionally i.i.d., so that if we define

$$a_n = a_n(F_n) = E\{ n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n) \mid F_n \}, \tag{4.12}$$

and

$$V_n^o = V_n^o(F_n) = E\{ n (\hat{\theta}_n^* - \hat{\theta}_n)^2 \mid F_n \}, \tag{4.13}$$

then, by (3.29) and the conditional i.i.d. nature of the estimators in (3.28), we conclude that given F_n ,

$$V_{nB}^* - V_n^o \rightarrow 0 \text{ a.s., as } M \rightarrow \infty . \tag{4.14}$$

However, a suitable rate of convergence in (4.14) may demand extra regularity conditions (particularly higher order monents of the $\hat{\theta}_n^*$). On the top of that , we may write

$$V_n^o = n.\mathrm{Var}\{ (\hat{\theta}_n^* - \hat{\theta}_n) \mid F_n \} + a_n^2 , \tag{4.15}$$

so that the behavior of V_n^o is not only dependent on the conditional variance in (4.15) but also on the asymptotic bias term in (4.12). In this context, we may note further that although $\psi(x,F(x))$ is twice differentiable with respect to x (as has been assumed earlier in this section) with $(d/dx)\psi(x,F(x)) = 0$ at $x = x_o$, the function $\psi(x,F_n(x))$ may not have a derivative (even first order) at $\hat{x}_{on}$. This may entail some non-smooth nature for which the rate of convergence for $\psi(x,F_n^*(x))$ may be somewhat slower. For example, in (4.1), if we replace $F(x_o)$ by $F_n(\hat{x}_{on})$) and want to have a similar identity, that won't work because F_n may not be differentiable at $\hat{x}_{on}$. As a result, in (4.3), with $F(x)$ replaced by $F_n(x)$, we may have a slower rate than $(x - \hat{x}_{on})^2 + [F_n(x) - F_n(\hat{x}_{on})]^2$. Moreover, even if $F(x) - F(y)$ is $O(n^{-1})$, $F_n(x) - F_n(y)$ may not be of the same order [Bahadur (1966) representation generally yields an order $n^{-3/4}(\log n)^{1/2}$ a.s.]. Similar slower rates of convergence remain applicable to the $F_n^*(x)$ and $\hat{x}_{on}^*$. Thus, we may not be in a position to claim the same rate of convergence as in (4.11).

We may remark in passing that the empirical distribution of the $n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n)$ in (3.28) may be used to provide a confidence interval for θ . But, any claim that such an interval behaves better than the conventional one based on (1.14) and (3.1) may demand the validity of the bootstrap Edgeworth expansion, which in this non-smooth case remains largely open for verification. The standard tools developed for functions of the mean type statistics may not be usable here.

REFERENCES

[1] BHATTACHARYYA, B.B. and SEN, P.K. (1976). Asymptotic normality of the extrema of certain sample functions. Zeit.Wahrsch.verw.Geb. 34, 113-118.

[2] DANIELS, H.A. (1945). The statistical theory of the strength of bundles of threads. Proc. Roy. Soc. Ser. A 183, 405-435.

[3] GANGOPADHYAY, A.K. and SEN, P.K. (1990). Bootstrap confidence intervals for conditional quantile functions. Sankhya, A 52, 346-363.

[4] GINE, E. and ZINN, J. (1990). Bootstrapping general empirical measures. Ann. Probability. 18, 851-869.

[5] SEN, P.K. (1973). On fixed-size confidence bands for the bundle strength of filaments. Ann. Statist. 1 , 526-537.

[6] SEN, P.K. (1976). An almost sure invariance principle for the extrema of certain sample functions. Ann. Probability 4, 80-89.

[7] SEN, P.K. (1981). Sequential Nonparametrics: Invariance Principles and Statistical Inference. John Wiley : New York.

[8] SEN, P.K. (1988a).Functional jackknifing: Rationality and general asymptotics. Ann. Statist. 16, 450-469.

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

[10] SEN, P.K. (1989). Whither delete-k jackknifing for smooth statistical functionals ? In Statistical Data Analysis and Inference (ed: Y. Dodge), North Holland: Amsterdam, pp. 269-279.

[11] SEN, P.K. (1990). Asymptotics via sub-sampling Bahadur-type representations. In Probability,Statistics and Design of Experiments: Essays in memory of R.C.Bose.(ed. R.R. Bahadur), Eastern Wiley: New Delhi, pp.653-666.

[12] SEN, P.K., BHATTACHARYYA, B.B. and SUH, M.W. (1973). Limiting behavior of the extrema of certain sample functions. Ann. Statist. 1, 297-311.