

SOME INFORMATIVE ASPECTS OF JACKKNIFING AND BOOTSTRAPPING

Pranab Kumar Sen

Departments of Biostatistics and Statistics

University of North Carolina

Chapel Hill, NC 27599-7400(3260), USA

The present study addresses a general query: Can an estimator be improved adaptively by a resampling plan? In this context, the conventional quadratic risk and the Pitman closeness criteria are incorporated in studying some informative aspects of the two principal resampling plans, namely, jackknifing and bootstrapping.

AMS Subject Classifications: 62G05, 62G99, 62F99

KEYWORDS AND PHRASES: Asymptotic bias; Asymptotic mean squared error; Asymptotic median unbiasedness; Pitman closeness; Quadratic risk; Shrinkage estimation; Statistical functionals.

1. INTRODUCTION

Let X_1, \dots, X_n be n independent and identically distributed random variables (i.i.d.r.v.) having a (cumulative) distribution function (d.f.) F , defined on the real line R . Let $\theta = \theta(F)$, a functional of the unknown d.f. F ,

be a parameter of interest, and based on the sample $\{X_1, \dots, X_n\}$, we want to draw statistical conclusions on $\theta(F)$. Let $F_n(x) = n^{-1} \sum_{i=1}^n I(X_i \leq x)$, $x \in \mathbb{R}$ be the empirical (sample) d.f. Then a natural (nonparametric) estimator of $\theta(F)$ is $T_n = \theta(F_n)$. Although, Sans the case of linear functions, T_n may not be unbiased for $\theta(F)$, under quite general regularity conditions, T_n is consistent, and as n increases,

$$n^{1/2}\{T_n - \theta(F)\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, v^2(F)), \quad (1.1)$$

where $v^2(F)$ (generally, a functional of the unknown d.f. F) is a nonnegative parameter. If the functional form of $v^2(F)$ is known and simple in nature, then $v_n^2 = v^2(F_n)$ can be taken as a consistent estimator of $v^2(F)$, so that for large values of n ,

$$n^{1/2}\{T_n - \theta(F)\}/v_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad (1.2)$$

This last result provides the key to drawing statistical conclusions on $\theta(F)$ for large n . Sans this simple situation, in many other cases, a resampling scheme may be incorporated in estimating $v^2(F)$ in a nonparametric fashion. Among these, jackknifing and bootstrapping methods are the most popular ones.

Let $F_{n-1}^{(i)}$ be the empirical d.f. of $\{X_1, \dots, X_n\} \setminus \{X_i\}$, and let

$$T_{n-1}^{(i)} = \theta(F_{n-1}^{(i)}) \quad \text{and} \quad T_{ni} = n T_n - (n-1)T_{n-1}^{(i)}, \quad i = 1, \dots, n; \quad (1.3)$$

$$T_{nJ} = n^{-1} \sum_{i=1}^n T_{ni} \quad \text{and} \quad v_{nJ}^2 = (n-1)^{-1} \sum_{i=1}^n (T_{ni} - T_{nJ})^2. \quad (1.4)$$

Then T_{nJ} is the classical jackknifed version of T_n and v_{nJ}^2 is the jackknifed (Tukey-) estimator of $v^2(F)$. If $E_F(T_n) = \theta(F) + n^{-1} \beta(F) + o(n^{-1})$, then $\beta(F)$ is taken as the asymptotic bias [of $n(T_n - \theta(F))$], and $E_F T_{nJ} = \theta(F) + o(n^{-1})$, so that jackknifing reduces the bias of an estimator. Under fairly general regularity conditions [viz., Sen (1988a,b)], as n increases,

$$(n-1)(T_{nJ} - T_n) + \beta(F) \xrightarrow{P} 0, \quad \text{and} \quad v_{nJ}^2/v^2(F) \xrightarrow{\text{a.s.}} v^2(F). \quad (1.5)$$

Thus, essentially, jackknifing leads to a bias reduction (up to the order n^{-1}) without compromising on the other properties of T_n and it provides with a nonparametric estimator of $v^2(F)$ as well.

Consider now M (usually large) independent samples, each of n independent observations drawn (with replacement) from the d.f. F_n . The observations in the i th sample are denoted by $X_{i1}^*, \dots, X_{in}^*$ and the corresponding empirical d.f. by $F_{n(i)}^*$, for $i=1, \dots, M$. Let then

$$T_{ni}^* = \theta(F_{n(i)}^*), \quad 1 \leq i \leq M; \quad T_{nB}^* = M^{-1} \sum_{i=1}^M T_{ni}^*; \quad (1.6)$$

$$v_{nB}^2 = (M-1)^{-1} \sum_{i=1}^M n(T_{ni}^* - T_{nB}^*)^2. \quad (1.7)$$

We term T_{nB}^* as a bootstrap version of T_n and v_{nB}^2 as bootstrap variance estimator. Then, under essentially the same regularity conditions as in the case of jackknifing, as $n \rightarrow \infty$, $M \rightarrow \infty$,

$$n^{1/2}(T_{nB}^* - T_n) \xrightarrow{P} 0 \quad \text{and} \quad v_{nB}^2/v^2(F) \xrightarrow{P} 1. \quad (1.8)$$

Moreover, if we consider the set $\{Y_{ni}^* = n^{1/2}(T_{ni}^* - T_n), \quad 1 \leq i \leq M\}$ and denote the empirical d.f. of the Y_{ni}^* by G_n^* (termed the bootstrap distribution), then under parallel regularity conditions,

$$G_n^* \xrightarrow{w} \Phi(\cdot; 0, v^2(F)), \quad \text{as } M \rightarrow \infty, n \rightarrow \infty, \quad (1.9)$$

where $\Phi(x; a, b^2)$ stands for the d.f. of a normal (a, b^2) r.v. [viz., Efron (1982)]. Although in the literature it has been emphasized that (1.9) may hold even if Φ were not a normal d.f., such a claim may not be universally true [viz., Athreya (1987)]. Further, for the computation of T_{nB}^* , v_{nB}^2 or G_n^* , one needs to generate nM ($= M^*$, say) i.i.d.r.v.'s from F_n . Although the advent of modern computers has made it possible to accomplish this task without much pain (even for large values of n and $M!$), there remains some arbitrariness in the proper choice of M (relative to n) as well as in the outcome (T_{nB}^*, v_{nB}^2) .

G_n^*). Two bootstrap samplers may generate different sets of bootstrap samples from the comma base sample (X_1, \dots, X_n) and thereby may end up with possible different values of T_{nB}^* , v_{nB}^2 and G_n^* . Also, the bootstrap samples are drawn from F_n (and not directly from F). Albeit F_n being sufficient for F (under general regularity conditions), the information contained in F_n can not presumably be equal to (or more than) that in F . Hence, we may pose a general query: Can T_{nB}^* (or T_{nJ}) be regarded as an improved version of T_n ? That is, is it possible to improve an estimator adaptively by a resampling plan? Incorporating the conventional quadratic risk and the Pitman closeness criteria, we intend to study some informative aspects of jackknifing and bootstrapping, the two popular resampling schemes.

2. SINGLE PARAMETER CASE

First, let us consider the case of the jackknifed version T_{nJ} in (1.4). it follows from Sen (1977) that

$$T_{nJ} = T_n + (n-1) E\{(T_n - T_{n-1}) | \mathcal{G}_n\} \quad (2.1)$$

where $\mathcal{G}_n = \mathcal{G}(X_{n:1}, \dots, X_{n:n}; X_{n+j}, j \geq 1)$ is the sigma-field generated by the order statistics $X_{n:1} \leq \dots \leq X_{n:n}$ and by $X_{n+j}, j \geq 1$, so that \mathcal{G}_n is \downarrow in n . Thus, T_{nJ} incorporates the bias correction through the conditional expectation in (2.1). If $\{T_n, \mathcal{G}_n; n \geq n_0\}$ forms a reversed martingale sequence (so that $E[T_{n-1} | \mathcal{G}_n] = T_n$ a.e., $\forall n \geq n_0$), as in the case with U-statistics, then $T_{nJ} = T_n$ with probability one, $\forall n > n_0$, and hence they are equally informative. The main role of jackknifing in this case relates to the nonparametric estimation of $v^2(F)$. As is typically the case with nonlinear functionals (such as the von Mises functions of degree ≥ 2) or nonlinear functions of reversed martingales, $\{T_n, \mathcal{G}_n; n \geq n_0\}$ may not strictly form a reversed

martingale sequence (and may not be unbiased too). In such a case, (2.1) provides an operational rule for bias reduction in addition to the variance estimation role of jackknifing. In the so called "smooth" case, we assume that

$$E_F T_n = \theta(F) + n^{-1} \beta(F) + o(n^{-1}), \quad (2.2)$$

so that under fairly general regularity conditions [viz., Sen (1988a,b)], (1.5) holds. In such a case, we have therefore,

$$\begin{aligned} & \text{Asymptotic mean square error (AMSE) of } n^{1/2}(T_n - \theta(F)) \\ &= \text{AMSE of } n^{1/2}(T_{nJ} - \theta(F)) + n^{-1} \beta^2(F) + o(n^{-1}), \end{aligned} \quad (2.3)$$

so that in the light of the usual quadratic risk criterion, T_{nJ} and T_n are asymptotically equivalent. T_{nJ} can still be regarded as an improved version of T_n in the sense that in (2.3), the second term on the right hand side is nonnegative, so that the bias correction in T_{nJ} has a variance reduction role (albeit very insignificant).

Next, we examine the picture in the light of the Pitman-closeness criterion (PCC). Recall that for two competing estimators, say $T_n^{(1)}$ and $T_n^{(2)}$, of a parameter θ , the PC measure may be defined as

$$PCM_{\theta}(T_n^{(1)}; T_n^{(2)}) = P_{\theta}\{|T_n^{(1)} - \theta| < |T_n^{(2)} - \theta|\} - P_{\theta}\{|T_n^{(1)} - \theta| > |T_n^{(2)} - \theta|\}, \quad (2.4)$$

and if (2.4) is nonnegative for all θ (and strictly positive for some θ), then $T_n^{(1)}$ is said to be 'closer' to θ than $T_n^{(2)}$, in the Pitman (1937) sense. If (2.4) converges to 0, uniformly in θ , $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are said to be asymptotically PC-equivalent. If for some positive integer k , (2.4) is $o(n^{-(k-1)/2})$, uniformly in θ (in a closed interval), $\{T_n^{(1)}\}$ and $\{T_n^{(2)}\}$ are said to be k -th order PC-equivalent. In the current context, $\theta = \theta(F)$, and hence, we shall conceive of a compact family \mathcal{F} of d.f.'s and allow F to be a member of \mathcal{F} . Thus, we have

$$\begin{aligned}
 \text{PCM}_F(T_{nJ}, F_n) &= P_F\{(T_{nJ}-T_n)^2 + 2(T_{nJ}-T_n)(T_n-\theta(F)) < 0\} \\
 &\quad - P_F\{(T_{nJ}-T_n)^2 + 2(T_{nJ}-T_n)(T_n-\theta(F)) > 0\} \\
 &= P_F\{(n-1)(T_{nJ}-T_n)(T_{nJ}+T_n-2\theta(F)) < 0\} \\
 &\quad - P_F\{(n-1)(T_{nJ}-T_n)(T_{nJ}+T_n-2\theta(F)) > 0\}. \tag{2.5}
 \end{aligned}$$

Note that by (1.2), $\{T_n\}$ is $\{n^{1/2}\}$ -consistent, so that if for a positive integer k ,

$$\limsup_{n \rightarrow \infty} n^{(k-1)/2} |P_F\{T_n \leq \theta(F)\} - \frac{1}{2}| = 0, \tag{2.6}$$

we term $\{T_n\}$ as k th order asymptotically median unbiased (AMU- k) for $\theta(F)$ [viz., Akahira and Takeuchi (1979)]. Note that by virtue of (1.2), $n^{1/2}\{\text{med}(T_n) - \theta(F)\} \rightarrow 0$ as $n \rightarrow \infty$ (a.e.F), so that (2.6) holds for some $k \geq 1$. On the other, if in (1.5), $\beta(F) \neq 0$, $n^{-1/2}\beta(F)$ contributes a term of the order $n^{-1/2}$ in $P\{T_{nJ}+T_n-2\theta(F) < 0\}$, so that even if $\{T_n\}$ is AMU-2, (2.5) is $O(n^{-1/2})$ (not $o(n^{-1/2})$). Thus, by virtue of (1.2), (1.5) and (2.6), we obtain that for $\beta(F) \neq 0$,

$$\text{PCM}_F(T_{nJ}; T_n) = \begin{cases} O(1), & \text{if } T_n \text{ is AMU-1} \\ O(n^{-1/2}), & \text{if } T_n \text{ is AMU-}k, k \geq 2. \end{cases} \tag{2.7}$$

If $\beta(F) = 0$, in (2.7), $O(n^{-1/2})$ can be replaced by $o(n^{-1/2})$ if T_n is AMU- k for some $k \geq 2$. Therefore, $\{T_n\}$ and $\{T_{nJ}\}$ are asymptotically first order PC-equivalent; for $\beta(F) = 0$ (a.e.F) and $\{T_n\}$ AMU- k for some $k \geq 2$, $\{T_n\}$ and $\{T_{nJ}\}$ are 2nd order PC-equivalent. In passing, we may remark that if $\beta(F) > 0$ and $\text{med}(T_n) \leq \theta(F) + \frac{1}{2n}\beta(F)$ (a.e.F), then (2.5) is ≥ 0 . Similarly, if $\beta(F) < 0$ and $\text{med}(T_n) \geq \theta(F) + \frac{1}{2n}\beta(F)$ (a.e.F), then (2.5) is ≥ 0 . Thus, the PC dominance of $\{T_{nJ}\}$ over $\{T_n\}$ may be judged by looking at $\beta(F)$ and the $\text{med}(T_n)$; but, this effect will be imperceptible for large n [by virtue of (2.7)].

Let us next consider a 'non-smooth' case where (2.2) may not hold, and the picture may be somewhat different. In the context of asymptotic normality and other related properties of $n^{1/2}(T_n - \theta(F))$, it is not uncommon to assume that T_n admits a first order asymptotic expansion [see Jurecková (1984)]:

$$T_n - \theta(F) = n^{-1} \sum_{i=1}^n \varphi_F(X_i) + R_n, \quad (2.8)$$

where the $\varphi_F(X_i)$ are i.i.d.r.v. with $E_F \varphi_F(X_i) = 0$,

$$0 < v^2(F) = E_F \varphi_F^2(X_i) < \infty \quad \text{and} \quad n^{1/2} R_n \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty. \quad (2.9)$$

Such a representation is actually justified by the classical Hoeffding-Hájek projection of T_n into independent summands, and in the literature, $\varphi_F(x)$ is known as the influence function of T_n (at x, F). Let us write $\bar{\varphi}_{F_n} = n^{-1} \sum_{i=1}^n \varphi_F(X_i)$, $n \geq 1$, and note that if T_n were a linear functional then $R_n \equiv 0$ with probability 1, and we would have

$$n(T_n - T_{n-1}) = n(T_n - \theta(F)) - n(T_{n-1} - \theta(F)) = \frac{n}{n-1} \{\varphi_F(X_n) - \bar{\varphi}_{F_n}\}, \quad n \geq 2. \quad (2.10)$$

Now, by the Khintchine strong law of large numbers, $\bar{\varphi}_{F_n} \rightarrow 0$ a.s., as $n \rightarrow \infty$,

while by (2.9), $n^{-1/2} \{ \max_{1 \leq i \leq n} |\varphi_F(X_i)| / v(F) \} \rightarrow 0$ a.s., as $n \rightarrow \infty$. Hence, from

(2.10), we have

$$n(T_n - T_{n-1}) - \{\varphi_F(X_n) - \bar{\varphi}_{F_n}\} = o(n^{-1/2}) \text{ a.s., as } n \rightarrow \infty. \quad (2.11)$$

Motivated by (2.11), we assume that $\{T_n\}$ admits a first order representation [as in (2.8)-(2.9)] and in addition, (2.11) holds. Note that (2.8) and (2.11) ensure that

$$n(R_n - R_{n-1}) = o(n^{-1/2}) \text{ a.s., as } n \rightarrow \infty, \quad (2.12)$$

and hence, using a version of the Hewitt-Savage Zero-One law and defining \mathcal{C}_n as in (2.1), we obtain that

$$E\{n(R_n - R_{n-1}) | \mathcal{C}_n\} = o(n^{-1/2}) \text{ a.s., as } n \rightarrow \infty. \quad (2.13)$$

As such, from (2.1), (2.8) and (2.13), we have under (2.11),

$$\begin{aligned}
 & (n-1) E[(T_n - T_{n-1}) | \mathcal{G}_n] \\
 &= n^{-1}(n-1) E\left[\frac{n}{n-1} (\varphi_F(X_n) - \bar{\varphi}_F(X_n) - \bar{\varphi}_n) | \mathcal{G}_n \right] + E\{n(R_n - R_{n-1}) | \mathcal{G}_n\} \\
 &= n^{-1}(n-1) E\{n(R_n - R_{n-1}) | \mathcal{G}_n\} \\
 &= o(n^{-1/2}) \quad \text{a.s., as } n \rightarrow \infty.
 \end{aligned} \tag{2.14}$$

Consequently, by (2.1) and (2.14), we have

$$n^{1/2}(T_{nJ} - T_n) \rightarrow 0 \quad \text{a.s., as } n \rightarrow \infty. \tag{2.15}$$

As such, parallel to (2.3), we have for large n ,

$$\text{AMSE of } n^{1/2}(T_n - \theta(F)) = \text{AMSE of } n^{1/2}(T_{nJ} - \theta(F)) + o(1), \tag{2.16}$$

and parallel to (2.7), we have

$$\text{PCM}_F(T_{nJ}; T_n) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \tag{2.17}$$

If we desire to improve the asymptotic equivalence results in (2.16) and (2.17), we may need suitable rate of convergence in (2.15), and this, in turn, may call for more refined results than in (2.11)-(2.12).

Let us now consider parallel results for the bootstrap estimator T_{nB}^* . Since for each $i (= 1, \dots, M)$, $T_{ni}^* = \theta(F_{ni}^*)$, under the usual regularity conditions on $\theta(\cdot)$ and $F(\Rightarrow F_n \Rightarrow F_{ni}^*)$, we may assume that given F_n , the T_{ni}^* admit a first order representation:

$$T_{ni}^* - T_n = n^{-1} \sum_{j=1}^n \varphi_{F_n}(X_{ij}^*) + R_{ni}^*, \quad 1 \leq i \leq M, \tag{2.18}$$

where the $\varphi_{F_n}(X_{ij}^*)$ are conditionally (given F_n) i.i.d.r.v.'s with $E_{F_n} \varphi_{F_n}(X_{ij}^*) = 0$ and $E_{F_n} \varphi_{F_n}^2(X_{ij}^*) = v_n^2(F_n)$, and further,

$$v_n^2(F_n) \rightarrow v^2(F), \quad \text{a.s./ist mean, as } n \rightarrow \infty. \tag{2.19}$$

Moreover, the R_{ni}^* are conditionally (given F_n) i.i.d.r.v., and

$$n E_{F_n} R_{ni}^{*2} \rightarrow 0, \quad \text{1st mean, as } n \rightarrow \infty. \quad (2.20)$$

Thus, on letting $R_{nB}^* = M^{-1} \sum_{i=1}^M R_{ni}^*$, we have

$$T_{nB}^* - T_n = (nM)^{-1} \sum_{i=1}^M \sum_{j=1}^n \varphi_{F_n}(X_{ij}^*) + R_{nB}^*, \quad (2.21)$$

where invoking the i.i.d. (given F_n) structure of the R_{ni}^* ,

$$n E_{F_n}(R_{nB}^{*2}) \leq n E(R_{n1}^{*2}) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (2.22)$$

From (2.21) and (2.22), we have

$$\begin{aligned} n E_{F_n}(T_{nB}^* - \theta(F))^2 &= n E_{F_n}[(T_{nB}^* - T_n) + (T_n - \theta(F))]^2 \\ &= n E_{F_n}\{E_{F_n}[(T_{nB}^* - T_n) + (T_n - \theta(F))]^2\} \\ &= M^{-1} E_{F_n}[v_n^2(F_n)] + n E_{F_n}(R_{nB}^{*2}) \\ &\quad + n E_{F_n}(T_n - \theta(F))^2 + 2 E_{F_n}[n R_{nB}^*(T_n - \theta(F))]. \end{aligned} \quad (2.23)$$

The first two terms on the right hand side of (2.23) are both nonnegative; the first term behaves like $M^{-1} v^2(F)$ while the second one converges to 0, as $n \rightarrow \infty$. Similarly, the last term (not affected by M) converges to 0 as $n \rightarrow \infty$. Actually, when R_{n1}^* has expectation (given F_n) equal to 0 (a.e. F_n), this last term is identically equal to 0. But, in (2.21), R_{nB}^* contains the conditional bias of T_{nB}^* (given F_n), and hence, in general, it may not have identically mean equal to 0. But, this conditional bias is generally $o(n^{-1/2})$, and this supports (2.20). Hence, from (2.23), we have

$$\begin{aligned} \text{AMSE of } n^{1/2}(T_{nB}^* - \theta(F)) &= \text{AMSE of } n^{1/2}(T_n - \theta(F)) \\ &\quad + M^{-1} v^2(F) + o(1), \quad \text{as } n \rightarrow \infty, \end{aligned} \quad (2.24)$$

so that the bootstrap version T_{nB}^* is asymptotically (as $n \rightarrow \infty$) equally informative as the original T_n only when M is chosen large. It is interesting to compare (2.3) and (2.24); whereas by jackknifing we are able to reduce the

AMSE (albeit insignificantly), by bootstrapping we have an increased value of the AMSE, unless M is very large. In fact, under the regularity conditions pertaining to (1.5), it can be shown that $n E(R_{nB}^{*2}) = O(\frac{1}{n})$, uniformly in M , so that in (2.24), $o(1)$ can be written as $O(n^{-1})$. Hence, in order to limit the excess of $AMSE [n^{\frac{1}{2}}(T_{nB}^* - \theta(F))]$ over $AMSE [n^{\frac{1}{2}}(T_n - \theta(F))]$ to $O(n^{-1})$ [as in (2.3)], we may need that $M=O(n)$. However, compared to (2.16), for (2.14), it suffices to take M large (without necessarily being $O(n)$). On the other hand, there is a basic concern with $M=O(n)$ when $\theta(\cdot)$ is not a linear functional. To illustrate this point, consider a 'smooth' functional $\theta(\cdot)$ admitting a second order expansion, for which in (2.18), we may write [as in Sen (1988a,b)]

$$R_{ni}^* = n^{-1} b_n(F_{ni}^*) + R_{ni}^{***}, \quad i \geq 1, \quad (2.25)$$

where $|b_n(F_{ni}^*) - b_n(F_n)| \rightarrow 0$ a.s., (a.e. F_n), as $n \rightarrow \infty$, while R_{ni}^{***} has conditional expectation, given F_n , $o(n^{-1})$ and $E(R_{ni}^{***2} | F_n) = O(n^{-2})$. We assume further that $b_n(F_n) \rightarrow \beta^*(F)$ a.s., as $n \rightarrow \infty$, where $|\beta^*(F)| < \infty$. If we use (2.25) in the definition of R_{nB}^* , we have

$$\begin{aligned} R_{nB}^* &= n^{-1} b_n(F_n) + o_p(n^{-1}) \\ &= n^{-1} \beta^*(F) + o_p(n^{-1}), \quad \text{as } n \rightarrow \infty. \end{aligned} \quad (2.26)$$

[Note that $\beta^*(F)$ may not be equal to $\beta(F)$ in (1.5).] Under appropriate regularity assumptions [on $b_n(F_{ni}^*)$], we may even replace $o_p(n^{-1})$ by $o(n^{-1})$ in second mean, so that for $M/n \rightarrow \gamma$ ($0 < \gamma < \infty$), we obtain that, given F_n ,

$$(nM)^{\frac{1}{2}}(T_{nB}^* - T_n) \xrightarrow{\mathcal{D}} N(\gamma^{\frac{1}{2}} \beta^*(F), v_n^2(F_n)), \quad (2.27)$$

and the AMSE of $n^{\frac{1}{2}}[T_n - \theta(F)]$ is given by

$$\begin{aligned} \text{AMSE of } n^{\frac{1}{2}}[T_n - \theta(F)] &+ \gamma^{-1} v^2(F) n^{-1} + n^{-1}(\beta^*(F))^2 \\ &+ n^{-1} 2 \beta^*(F)\beta(F) + o(n^{-1}), \end{aligned} \quad (2.28)$$

where in (2.28) we have made use of (2.23), (2.26) and (2.2) for $T_n - \theta$. Note

that $\beta^*(F)$ may be different from 0 even if $\beta(F) = 0$, and, in general, $\beta^*(F)$ and $\beta(F)$ may not be equal or of the same sign. However, noting that AMSE of $n^{1/2}[T_n - \theta(F)] = v^2(F) + n^{-1} \beta^2(F) + o(n^{-1})$, we obtain that (2.28) can be written as

$$v^2(F) + n^{-1} \{ \gamma^{-1} v^2(F) + (\beta^*(F) + \beta(F))^2 \} + o(n^{-1}). \quad (2.29)$$

It may be recalled that $\beta(F)$ arises mainly due to the expected value of the second order compact derivative of $\theta(\cdot)$ at F , and, similarly, $b_n(F_n)$ comes from the conditional expectation (given F_n) of the second order compact derivative of $\theta(\cdot)$ at F_n , when we assume that $\theta(\cdot)$ is second order Hadamard differentiable in a neighborhood of F (containing the F_n a.s., as $n \rightarrow \infty$). Since typically this second order compact derivative is not a constant nor a linear functional, we may not have

$$E_F b_n(F_n) = b_n(E_F F_n) = b_n(F), \quad (2.30)$$

and hence, the convergence of $b_n(\cdot)$ to some $b(\cdot)$ may not necessarily imply that $E_F b_n(F_n) \rightarrow b(F)$. Now (2.29) clearly illustrates the role of the asymptotic bias terms $\beta(F)$ and $\beta^*(F)$ and $\gamma (= M/n)$ in the inflation of the AMSE of the bootstrap estimator. In some simple situations, $\beta(F) = \beta^*(F)$ and (2.29) indicates that bootstrapping incorporates a four times effect of the bias square in the second order term of its AMSE as well as a factor $O(M^{-1})$ due to bootstrapping. As an illustration, consider the simple functional:

$$\begin{aligned} \theta(F) &= \frac{1}{2} \iint (x_1 - x_2)^2 dF(x_1) dF(x_2) \\ &= \text{Variance of } X. \end{aligned} \quad (2.31)$$

Here $\beta(F) = -\theta(F)$, $\theta(F_n) = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = s_n^2$ and $b_n(F_n) = -s_n^2 \rightarrow -\theta(F)$ a.s., as $n \rightarrow \infty$, so that $\beta^*(F) = \beta(F) = -\theta(F)$. Thus, using the biased estimator (s_n^2) of $\theta(F)$, bootstrapping has four fold effect of $(\text{bias})^2$ in its AMSE. If we use $T_n = n(n-1)^{-1} s_n^2$, the unbiased estimator of $\theta(F)$, we would

have $\beta^*(F) = \beta(F) = 0$; nevertheless, $M^{-1} v^2(F)$ reflects the excess due to bootstrapping. If the regularity conditions pertaining to (2.26) do not hold [viz., $\theta(F) = F^{-1}(p)$ for some $p \in (0,1)$], (2.28) may involve other terms $O(n^{-\lambda})$ for some $1 > \lambda > 0$, and hence, the excess may become $O(n^{-\lambda})$, for some $\lambda \in (0,1)$, even when we choose $M = O(n)$. Hence, if λ is known, M may as well be chosen as $O(n^\lambda)$, and there is no incentive in choosing M unnecessarily large. The picture becomes somewhat diffused for moderate values of n , and hence, there is a good deal of interest in the moderate sample asymptotics in this context.

Let us next consider the PC-performance of the bootstrap version T_{nB}^* relative to T_n . Writing $T_{nB}^* - \theta(F) = (T_{nB}^* - T_n) + (T_n - \theta(F))$ and proceeding as in (2.5), we obtain that

$$\begin{aligned} \text{PCM}_F(T_{nB}^*; T_n) &= P_F\{(T_{nB}^* - T_n)^2 + 2(T_{nB}^* - T_n)(T_n - \theta(F)) < 0\} \\ &\quad - P_F\{(T_{nB}^* - T_n)^2 + 2(T_{nB}^* - T_n)(T_n - \theta(F)) > 0\} \\ &= E_F[P_F\{(T_{nB}^* - T_n)^2 + 2(T_{nB}^* - T_n)(T_n - \theta(F)) < 0\}] \\ &\quad - E_F[P_F\{(T_{nB}^* - T_n)^2 + 2(T_{nB}^* - T_n)(T_n - \theta(F)) > 0\}]. \end{aligned} \quad (2.32)$$

First consider the "smooth" case, where under appropriate regularity conditions, we invoke (2.27), (1.1), (2.2) and assume that $M/n \rightarrow \gamma: 0 < \gamma < \infty$.

We write

$$\begin{aligned} (T_{nB}^* - T_n)^2 + 2(T_{nB}^* - T_n)(T_n - \theta) &\lesssim 0 \\ \Leftrightarrow \frac{1}{\sqrt{M}} \{\sqrt{Mn}(T_{nB}^* - T_n)\}^2 + 2\sqrt{nM}(T_{nB}^* - T_n)\sqrt{n}(T_n - \theta(T)) &\lesssim 0, \end{aligned} \quad (2.33)$$

where by (2.27), $\{\sqrt{Mn}(T_{nB}^* - T_n)\}^2 = O_p(1)$, and by (1.1), T_n is AMU-1, so that by (2.33) and some routine steps, (2.32) converges to 0, as n (or M) $\rightarrow \infty$. Thus,

$$\text{PCM}_F(T_{nB}^*; T_n) \rightarrow 0, \text{ as } n \rightarrow \infty, \quad (2.34)$$

and hence, $\{T_{nB}^*\}$ and $\{T_n\}$ are asymptotically PC-equivalent. Note that if T_n is AMU-2 for $\theta(F)$, then by (2.27) and (2.33), (2.32) is $O(M^{-1/2}) = O(n^{-1/2})$, so that the picture is quite comparable to (2.7). Actually, for (2.34) to hold, we do not need (2.27). Recall that $\sqrt{n}(T_{nB}^* - T_n) \xrightarrow{P} 0$ as $n \rightarrow \infty$, and hence, in (2.33), dropping M on the right hand side, we still claim that (2.34) holds. Thus, for the asymptotic PC-equivalence result in (2.34), (1.1) and (1.8) suffice. But, to obtain an $O(n^{-1/2})$ rate for (2.34), conditional AMU- k property of T_{nB}^* (given F_n), for $k \geq 2$, may require more elaborate analysis than in the case of T_n .

Let us compare T_{nJ} and T_{nB}^* in the light of the PCM. Unfortunately, due to possible lack of transitivity [c.f. Blyth (1972)], (2.7) and (2.34) may not by themselves convey the entire picture. However, we may proceed as in (2.32) and write

$$\begin{aligned} \text{PCM}_F(T_{nB}^*; T_{nJ}) &= E_F [P_{F_n} \{ (T_{nB}^* - T_{nJ})^2 + 2(T_{nB}^* - T_{nJ})(T_{nJ} - \theta(F)) < 0 \} \\ &\quad - P_{F_n} \{ (T_{nB}^* - T_{nJ})^2 + 2(T_{nB}^* - T_{nJ})(T_{nJ} - \theta(F)) > 0 \}]. \end{aligned} \quad (2.35)$$

If T_{nJ} is median unbiased (MU) for $\theta(F)$ while conditionally, given F_n , T_{nB}^* is MU for T_{nJ} , then (2.35) is ≤ 0 , so that in the light of the PCM, T_{nJ} is closer to $\theta(F)$ than T_{nB}^* . In a variety of situations, particularly, in the location-scale models, T_{nJ} ($\equiv T_n$) has a d.f. symmetric about $\theta(F)$, so that T_{nJ} is MU for $\theta(F)$. In scale models, T_n may be so adjusted by a scalar factor that T_{nJ} becomes MU for $\theta(F)$, and the choice of such a jackknifed version may be made on the ground of equivariance, as has been elaborated in Ghosh and Sen (1989) and Nayak (1990). On the otherhand, F_n is a discrete d.f. with probability mass n^{-1} at the given points x_1, \dots, x_n , and hence, the conditional distribution of T_{nB}^* , given F_n , may not be symmetric about T_n (or T_{nJ}) or may

not have even median equal to T_{nJ} . On the set (F_n) where the conditional median of T_{nB}^* , given F_n , is nonnegative, (2.35) is ≤ 0 , while on the complementary set, (2.35) may be both \leq or ≥ 0 , depending on dispersion of $T_{nB}^* - T_{nJ}$. Invoking symmetry of F and on the functional $\theta(\cdot)$, as is the case in R-, M- and L- estimation of location, we may argue that the part of the sample space (in R^n) for which the conditional median of $T_{nB}^* - T_{nJ}$, given F_n , is ≤ 0 has probability 1/2, and hence, in such a case, (2.35) is ≤ 0 , so that T_{nJ} dominates T_{nB}^* is the PCC. This argument may run into difficulties in a general case where such symmetric structures are not easy to comprehend. Nevertheless, we may argue as in after (2.32) and show that

$$\text{PCM}_F(T_{nB}^*, T_{nJ}) \rightarrow 0 \text{ as } n \rightarrow \infty \text{ (} M \rightarrow \infty \text{)}, \quad (2.36)$$

while in the "smooth" case, (2.36) can be strengthened to $O(n^{-1/2})$ when $M = O(n)$ and both T_{nJ} and T_{nB}^* are AMU- k for some $k \geq 2$. In the small to moderate sample cases, by a skillful choice of T_{nJ} , we expect a better PC-performance than T_{nB}^* , although for large values of n (and M), this effect is imperceptible. In the formulation of the bootstrap estimator T_{nB}^* , we have adopted the so called 'equal probability sampling' schemes, so that the T_{ni}^* , $i \geq 1$, are conditionally, given F_n , i.i.d.r.v.'s. There are some other variants of bootstrap sampling schemes primarily designed to reduce M subject to a level of accuracy of the bootstrap estimators comparable to the equal probability sampling scheme; balanced sampling, importance sampling and a few other schemes have been incorporated in this context. We may refer to Hall (1989 a,b) where other references have been cited. So long as (2.18) is justifiable for the individual bootstrap sample estimates, we may as well derive a variant form of (2.21), and thereby verify that the conclusions made before all pertain to such schemes, allowing M to be comparatively small. Nevertheless, one needs to assume that $M \rightarrow \infty$ as $n \rightarrow \infty$, although the rate is

generally somewhat slower. Judged from the PCM point of view, it seems more pertinent to choose $MU T_{nJ}$ and to adopt a bootstrap samplign scheme such that the conditional median of T_{nB}^* , given F_n , is $\leq T_{nJ}$ more often than being $> T_{nJ}$. That would bring (2.35) close to 0 and make T_{nB}^* quite favorably comparable to the T_{nJ} .

3. MULTI-PARAMETER CASE

Consider a generalization of the model in Section 1 where

$$\underline{\theta} = \underline{\theta}(F) = (\theta_1(F), \dots, \theta_p(F))' \quad (3.1)$$

is a p -vector ($p \geq 1$) of functionals of the underlying d.f. F , and the d.f. F may as well be a multivariate one; the sample observations are denoted by X_1, \dots, X_n . We denote the empirical d.f. by F_n and consider the estimator (vector)

$$\underline{T}_n = (T_{n1}, \dots, T_{np})' = (\theta_1(F_n), \dots, \theta_p(F_n))'. \quad (3.2)$$

As a natural extension of (1.1), we have under appropriate regularity conditions,

$$n^{1/2}(\underline{T}_n - \underline{\theta}) \xrightarrow{\mathcal{D}} \mathcal{N}_p(Q, \underline{\nu}(F)), \text{ as } n \rightarrow \infty, \quad (3.3)$$

where the nonnegative definite (nnd) matrix $\underline{\nu}(F)$, a matrix-valued functionals of F , may again be consistently estimated by jackknifing or bootstrapping methods; we denote these estimators by \underline{V}_{nJ} and \underline{V}_{nB}^* respectively.

For the jackknifed version of \underline{T}_n , we define the pseudevariables (vectors) \underline{T}_{ni} as in (1.3) and let (as in (1.4))

$$\underline{T}_{nJ} = n^{-1} \sum_{i=1}^n \underline{T}_{ni}; \quad \underline{V}_{nJ} = (n-1)^{-1} \sum_{i=1}^n (\underline{T}_{ni} - \underline{T}_{nJ})(\underline{T}_{ni} - \underline{T}_{nJ})'. \quad (3.4)$$

For the bootstrapped version of \underline{T}_n , we define the \underline{T}_{ni}^* as in (1.6) and let

$$\tilde{T}_{nB}^* = M^{-1} \sum_{i=1}^M \tilde{T}_{ni}^*; \tilde{V}_{nB}^* = (M-1)^{-1} \sum_{i=1}^M (\tilde{T}_{ni} - \tilde{T}_{nB}^*)(\tilde{T}_{ni} - \tilde{T}_{nB}^*)'. \quad (3.5)$$

In the conventional case, we may consider a quadratic loss

$$L_n(\tilde{T}, \tilde{\theta}) = n(\tilde{T} - \tilde{\theta})' Q(\tilde{T} - \tilde{\theta})$$

where Q is a given nnd matrix, and define $\rho_n(\tilde{\theta}) = E L_n(\tilde{T}_n, \tilde{\theta})$,

$$\rho_{nJ}(\tilde{\theta}) = E L_n(\tilde{T}_{nJ}, \tilde{\theta}), \quad (3.7)$$

$$\rho_{nB}(\tilde{\theta}) = E L(\tilde{T}_{nB}^*, \tilde{\theta}). \quad (3.8)$$

Then, we may compare $\rho_{nJ}(\tilde{\theta})$ and $\rho_{nB}(\tilde{\theta})$ to draw conclusions about the relative performance of the jackknifed and bootstrapped version of \tilde{T}_n .

In the "smooth" case (where (2.2) holds in a vector-form), we have parallel to (2.3), as $n \rightarrow \infty$,

$$\rho_n(\tilde{\theta}) = \rho_{nJ}(\tilde{\theta}) + n^{-1} [\beta(F)]' Q[\beta(F)] + o(n^{-1}), \quad (3.9)$$

so that the jackknifed version has a slight advantage over the classical version, and this is mainly due to the "bias" reduction role of jackknifing. Similarly, parallel to (2.24), we have

$$\rho_{nB}(\tilde{\theta}) = \rho_n(\tilde{\theta}) + M^{-1} \text{Tr}(Qv) + o(1), \quad \text{as } M \rightarrow \infty, n \rightarrow \infty, \quad (3.10)$$

and in (3.10), $o(1)$ can be replaced by $O(M^{-1})$ if $M = O(n)$ and other smoothness conditions discussed after (2.24) hold in the vector case too. From (3.9) and (3.10), we may conclude that jackknifing may be generally (slightly) more informative than bootstrapping.

Let us examine the relative picture in the light of the POC (which we need to extend to the vector case too). With the definition of the loss function $L(\cdot)$ in (3.6), we extend (2.4) as

$$\text{GPC } M_{\tilde{\theta}}(\tilde{T}_n^{(1)}; \tilde{T}_n^{(2)}) = P_{\tilde{\theta}}\{L(\tilde{T}_n^{(1)}, \tilde{\theta}) < L(\tilde{T}_n^{(2)}, \tilde{\theta})\} - P_{\tilde{\theta}}\{L(\tilde{T}_n^{(1)}, \tilde{\theta}) > L(\tilde{T}_n^{(2)}, \tilde{\theta})\} \quad (3.11)$$

As such, as in (2.5), we have

$$\begin{aligned} \text{GPC } M_F(\tilde{T}_{nJ}; \tilde{T}_n) &= P_F\{(n-1)(\tilde{T}_{nJ} - \tilde{T}_n)' Q(\tilde{T}_{nJ} + \tilde{T}_n - 2\theta(F)) < 0\} \\ &\quad - P_F\{(n-1)(\tilde{T}_{nJ} - \tilde{T}_n)' Q(\tilde{T}_{nJ} + \tilde{T}_n - 2\theta(F)) > 0\}. \end{aligned} \quad (3.12)$$

As in Sen (1991), we introduce the notion of multivariate median unbiasedness (MMU) property and extend (2.6) as follows:

If for a positive integer k and a $n^{1/2}$ -consistent estimator $\{\tilde{T}_n\}$,

$$\limsup_{n \rightarrow \infty} \sup_{F \in \mathcal{F}} \sup_{\underline{a} \in U} n^{(k-1)/2} |P_F\{\underline{a}'(\tilde{T}_n - \theta(F)) \leq 0\} - \frac{1}{2}| = 0, \quad (3.13)$$

where $U = \{\underline{a} : \|\underline{a}\|^2 = 1\}$, then $\{\tilde{T}_n\}$ is k th order asymptotically multivariate median unbiased (AMMU- k) for $\theta(F)$. Diagonal symmetry (or nearly so) of the d.f. of \tilde{T}_n around $\theta(F)$ ensures the MMU-property. In the smooth case, taking into account (1.5) (coordinatewise), we see that whenever \tilde{T}_n is AMMU- k , for some $k \geq 1$, (3.12) is $o(1)$, so that \tilde{T}_{nJ} and \tilde{T}_n are asymptotically GPC-equivalent. An order of $n^{-1/2}$ may also be obtained when (3.13) holds for some $k \geq 2$. Also, for (3.12) to be $o(1)$, it is not necessary to be confined only to the smooth case. If we assume that the first order representation in (2.8) along with (2.11) hold coordinatewise, then (2.17) directly extends to $\text{GPCM}_F(\tilde{T}_{nJ}; \tilde{T}_n) \rightarrow 0$ as $n \rightarrow \infty$. The treatment of asymptotic GPC-equivalence of \tilde{T}_{nB}^* and \tilde{T}_n runs parallel to that in (2.32) through (2.34) with direct adaptations from (3.12) and (3.13). Hence, omitting these steps, we claim that $\text{GPCM}_F(\tilde{T}_{nB}^*; \tilde{T}_n) \rightarrow 0$ as $n \rightarrow \infty$. Finally, (2.35) also extends directly to the vector case, and hence, if \tilde{T}_{nJ} is MMU for $\theta(F)$, it may have an advantage over the bootstrap estimator \tilde{T}_{nB}^* , when n is not so large.

In the multiparameter case, \tilde{T}_n may not necessarily be optimal (even asymptotically) although in a coordinatewise setup an (asymptotic) optimality property can be established. In the literature, this effect is known as the

Stein-effect, and there exist some shrinkage estimators (also known as Stein-rule estimators) which dominate \tilde{T}_n in quadratic risk (as well as in other criteria). Stein-rule versions are typically non-linear having a dominant "pre-test" flavor and are generally biased. The introduction of the bias term is deliberate with the objective of lower the risk. On the other hand, jackknifing is primarily incorporated to reduce the 'bias' without compromising on the dispersion of an estimator. As such, it seems counter intuitive to use jackknifing on Stein-rule estimators. In fact, in some simple multivariate model, it has been shown [viz., Sen (1986)] that jackknifing a Stein-rule estimator may not reduce the risk any further (and, rather, make it comparatively larger) and may fail to estimate the risk also. If we assume that coordinatewise \tilde{T}_n satisfies the first order representation in (2.8) [along with (2.11)], then similar conclusions hold for jackknifed Stein-rule versions of \tilde{T}_n . As such, there is not much incentive in examining the informative aspects of such estimators. The situation may be somewhat different with the bootstrapping, and hence, we shall discuss some Stein-rule bootstrap versions of \tilde{T}_n for depicting this situation.

Recall that the bootstrap version of \tilde{T}_n considered here is

$$\tilde{T}_{nB}^* = M^{-1} \sum_{i=1}^M \tilde{T}_{ni}^* \tag{3.14}$$

where the \tilde{T}_{ni}^* are i.i.d.r. vectors (conditionally on F_n), so that \tilde{T}_{nB}^* really estimates \tilde{T}_n . Hence, it may be of some interest to inquire whether shrinking \tilde{T}_{nB}^* towards \tilde{T}_n makes it more informative (on $\vartheta(F)$)? As such, we consider the following bootstrap Stein-rule estimator:

$$\tilde{T}_{nB}^{*s} = \tilde{T}_n + \{I - c_n d_n \varrho_n^{*-1} Q^{-1} \varrho_n^{-1}\} (\tilde{T}_{nB}^* - \tilde{T}_n), \tag{3.15}$$

where c_n is a positive number (converging to a limit $c > 0$), $d_n = \text{ch}_p(Q \varrho_n)$ is the smallest characteristic root of $Q \varrho_n$, ϱ_n is defined by (3.4) and

$$\varphi_n^* = nM(\tilde{T}_{nB}^* - \tilde{T}_n)' \tilde{V}_{nJ}^{-1}(\tilde{T}_{nB}^* - \tilde{T}_n). \quad (3.16)$$

It is also possible to replace \tilde{V}_{nJ} by the bootstrap version \tilde{V}_{nB}^* , defined by (3.5). But conditional on F_n , \tilde{V}_{nB}^* behaves like a stochastic matrix, converging stochastically to \tilde{V}_{nJ} , while \tilde{V}_{nJ} is held fixed under the bootstrap sampling scheme. Hence, it is more natural to consider \tilde{V}_{nJ} instead of \tilde{V}_{nB}^* in (3.15) and (3.16). Recall that by (3.15),

$$\tilde{T}_{nB}^{*s} - \varrho(F) = [\tilde{T}_n - \varrho(F)] + \{\tilde{I} - c_n d_n \varphi_n^{*-1} \tilde{Q}^{-1} \tilde{V}_{nJ}^{-1}\}(\tilde{T}_{nB}^* - \tilde{T}_n) \quad (3.17)$$

so that we have

$$\begin{aligned} E_F\{(\tilde{T}_{nB}^{*s} - \varrho(F))' \tilde{Q}(\tilde{T}_{nB}^{*s} - \varrho(F))\} &= E_F[(\tilde{T}_n - \varrho(F))' \tilde{Q}(\tilde{T}_n - \varrho(F))] \\ &+ 2 E_F\{(\tilde{T}_n - \varrho(F))' \tilde{Q} E_F[\{\tilde{I} - c_n d_n \varphi_n^{*-1} \tilde{Q}^{-1} \tilde{V}_{nJ}^{-1}\}(\tilde{T}_{nB}^* - \tilde{T}_n)]\} \\ &+ E_F\{(\tilde{T}_{nB}^* - \tilde{T}_n)' \{\tilde{I} - c_n d_n \varphi_n^{*-1} \tilde{Q}^{-1} \tilde{V}_{nJ}^{-1}\}' \tilde{Q} \{\tilde{I} - c_n d_n \varphi_n^{*-1} \tilde{Q}^{-1} \tilde{V}_{nJ}^{-1}\}(\tilde{T}_{nB}^* - \tilde{T}_n)\}. \end{aligned} \quad (3.18)$$

The last term on the right hand side of (3.18) is nonnegative and is of the order $(nM)^{-1}$. For linear functional, $E_f \{\tilde{T}_{nB}^* - \tilde{T}_n\} = \tilde{Q}$, and in our setup, $E_{F_n} \{\tilde{T}_{nB}^* - \tilde{T}_n\} = o_p(n^{-1/2})$, while it can be $O_p(n^{-1})$ under additional regularity conditions [including $M = O(n)$]. Also, in the contemplated bootstrap sampling scheme, given F_n , as n increases,

$$\varphi_n \xrightarrow{\mathcal{D}} \chi_p^2, \text{ a central chi square r.v. with PDF.} \quad (3.19)$$

As such, we may use a variant of the Stein identity in a conditional (but asymptotic) setup, and claim that as n increases,

$$E_{F_n} \{[\tilde{I} - c_n d_n \varphi_n^{*-1} \tilde{Q}^{-1} \tilde{V}_{nJ}^{-1}] (\tilde{T}_{nB}^* - \tilde{T}_n)\} = o_p(n^{-1/2}), \quad (3.20)$$

while it can as well be equal to \tilde{Q} (a.e.) for linear functionals or $O(n^{-1})$ under additional regularity conditions. Since $n^{1/2}(\tilde{T}_n - \varrho(F))$ is asymptotically normal, it seems that the second term on the right hand side of (3.18) can be

made $o(n^{-1})$ in general, and $O(n^{-3/2})$ under additional regularity conditions; it may also be equal to 0 in some special cases. Thus,

$$\begin{aligned} n E_F \{ (T_{nB}^{*s} - \theta(F))' Q(T_{nB}^{*s} - \theta(F)) \} &= n E_F \{ (T_n - \theta(F))' Q(T_n - \theta(F)) \} \\ &+ o(1) + O(M^{-1}), \quad \text{as } n \rightarrow \infty. \end{aligned} \quad (3.21)$$

Consequently, shrinking T_{nB}^* towards T_n may not provide any extra incentive unless M is large. Even so, $T_{nB}^* - T_n = o_p((nM)^{-1/2})$, and hence, such a shrinking is of any perceptible effect. A very similar picture holds for the GPC-criterion.

Next, we may enquire whether shrinking a bootstrap estimator towards an arbitrary pivot θ_0 makes it more appealing than the parallel shrinkage version of T_n ? Without any loss of generality, we let $\theta_0 = \theta$. Consider then

$$T_n^S = \{ I - c_n d_n \varphi_{n0}^{-1} Q^{-1} V_{nJ}^{-1} \} T_n, \quad (3.22)$$

$$T_n^{S*} = \{ I - c_n d_n \varphi_{n*}^{-1} Q^{-1} V_{nJ}^{-1} \} T_{nB}^*$$

(3.23)

where $\varphi_{n0} = n(T_n)' V_{nJ}^{-1}(T_n)$ and $\varphi_{n*} = nM(T_{nB}^*)' V_{nJ}^{-1}(T_{nB}^*)$. Again, we may write

$$T_n^{S*} - \theta(F) = (T_{nB}^* - T_n) + (T_n^S - \theta(F)) + U_n, \quad (3.24)$$

where

$$U_n = c_n d_n Q^{-1} V_{nJ}^{-1} \{ (\varphi_{n0}^{-1} - \varphi_{n*}^{-1}) T_n + \varphi_{n*}^{-1} (T_n - T_{nB}^*) \}. \quad (3.25)$$

In passing, we may remark that in (3.16), conditional on F_n , φ_n^* has asymptotically (in probability) central chi square distribution with p DF, when actually $(nM)^{1/2}$. $(T_{nB}^* - T_n)$ has asymptotically $N_p(0, V_n)$, conditionally on F_n . For the last result, we need to presume that the asymptotic bias of $(nM)^{1/2} (T_{nB}^* - T_n)$, conditional on F_n , is negligible, i.e.,

$$E\{ (T_{nB}^* - T_n) | F_n \} = o_p((nM)^{-1/2}). \quad (3.26)$$

But $E(T_{nB}^{*} | F_n) = E(T_{ni}^{*} | F_n)$, $\forall M$, and hence, (3.26) may also be written as

$$E\{(T_{n1}^{*} - T_n) | F_n\} = o_p((nM)^{-1/2}), \text{ as } n \rightarrow \infty. \quad (3.27)$$

In particular, if we let $M = O(n)$, then (3.27) leads to $o(n^{-1})$. Let us also note that

$$E\{\varphi_{n*}^{-1} | F_n\} \sim E\{\chi_{p, M\varphi_{no}}^{-2}\}, \quad (3.28)$$

so that $E(\varphi_{no}^{-1} - \varphi_{n*}^{-1}) \sim E\{\varphi_{no}^{-1} - \chi_{p, M\varphi_{no}}^{-2}\}$ which is a nonnegative quantity. Thus,

operating with the loss function in (3.6), using (3.24) and incorporating a two-stage expectation (first under F_n and then over F_n), we obtain that

$$E_F L_n(T_n^{S*}, \theta(F)) = E_F L_n(T_n^S, \theta(F)) + M^{-1} \text{Tr}(Qv) + o(1), \quad (3.29)$$

so that we have a similar conclusion that for large M , granted (3.27), T_n^{S*} and T_n^S are asymptotically (as $n \rightarrow \infty$) equivalent, while for finite M , the shrinkage bootstrap estimator T_n^{S*} is less informative than T_n^S . A similar result holds if we use (3.24) and the GPC-criterion. In passing, we may remark that in an asymptotic setup, the shrinkage estimator T_n^S behaves better than T_n only when $\theta(F)$ lies in a Pitman-neighborhood of θ_0 i.e., $n^{1/2} \|\theta(F) - \theta_0\| = O(1)$; otherwise, $n^{1/2} \|T_n^S - T_n\| \xrightarrow{P} 0$, as $n \rightarrow \infty$. As such, we need to confine ourselves to a Pitman-neighborhood of the pivot $\theta_0 = (= \underline{0})$ when studying (3.29), and even so, the second term on the right hand side clearly reflects the dominant role M plays in this context.

4. SOME GENERAL REMARKS

In Sections 2 and 3, we have tacitly assume that T_n is closely approximable by a linear functional for which the asymptotic normality result holds under appropriate regularity conditions. There may be some situations

where such a linear functional is degenerate, so that $n^{1/2}(T_n - \theta(F))$ has asymptotically a degenerate distribution. In such a case, (1.9) is redundant. Much of the charms of bootstrapping may also vanish in such a non-regular case; Athreya's (1987) findings are pertinent in this context. In this setup, the use of a quadratic risk function may be dubious, and a generalized PC-criterion may provide a better interpretation. In such a non-regular case, the variance-estimation role of jackknifing may also become questionable. However, the bias reduction role of jackknifing remains as a dominant one, especially because the bias of T_n and its asymptotic standard error may be of the same order. Thus, from this point of view too, jackknifing may have some advantages over bootstrapping.

Let us consider some other situations which may arise when T_n is linearly approximable, but not too smoothly. As an example, consider the 'bundle strength of filaments' model [Sen (1973)] where X_1, \dots, X_n are i.i.d.r.v.'s with a d.f. F defined on R^+ , and where

$$T_n = n^{-1} \left\{ \max_{1 \leq i \leq n} (n-i+1)X_{n:i} \right\} \quad (4.1)$$

and the $X_{n:i}$ are the order statistics corresponding to the X_i . We may rewrite T_n as

$$T_n = \sup \{x[1-F_n(x-)] : x \in R^+\} \quad (4.2)$$

and define

$$\theta(F) = \sup \{x[1-F(x)] : x \in R^+\}. \quad (4.3)$$

Note that $\{T_n; n \geq 1\}$ forms a reverse sub-martingale and if x_0 is the unique point such that $x_0(1-F(x_0)) = \theta(F)$, then on letting

$$M_n = x_0[F_n(x_0) - F(x_0)] = n^{-1} \sum_{i=1}^n \{x_0[I(X_i \leq x_0) - F(x_0)]\}, \quad (4.4)$$

$$n^{1/2}[(T_n - \theta(F)) + M_n] \rightarrow 0 \text{ a.s., as } n \rightarrow \infty. \quad (4.5)$$

It has also been shown that $n^{1/2}|E(T_n) - \theta(F)| \rightarrow 0$ as $n \rightarrow \infty$, although $E(T_n) -$

$\theta(F)$ may not be $O(n^{-1})$. The asymptotic normality result for $n^{1/2}(T_n - \theta(F))$ follows directly from (4.4) and (4.5), although jackknifing may not eliminate the bias up to the order n^{-1} . Motivated by this example, we consider the case where for a functional $T_n = \theta(F_n)$,

$$E_F(T_n) = \theta(F) + n^{-\lambda} a(F) + o(n^{-\lambda}), \text{ for some } \lambda \in (1/2, 1]. \quad (4.6)$$

In this case, we have

$$E_F(T_{nJ}) = \theta(F) + (1-\lambda)n^{-\lambda} a(F) + o(n^{-\lambda}). \quad (4.7)$$

As such, we do not expect (1.5) to hold; rather, we would have

$$n^\lambda(T_n - T_{nJ}) + \lambda a(F) \xrightarrow{P} 0, \text{ as } n \rightarrow \infty. \quad (4.8)$$

This result may play an intricate role in the ascertainment of the AMU-k property in (2.6), and may lead to the supremacy of one of T_n and T_{nJ} over the other in the PC sense; in quadratic risk sense, we would have in (2.3), $n^{-1} \beta^2(F)$ replaced by $n^{1-2\lambda} \lambda^2 a^2(F)$. Thus, for $\lambda < 1$, we have a smaller discrepancy.

It is not necessary to assume that $T_n (= \theta(F_n))$ are functionals of F_n alone. We may as well consider arbitrary T_n satisfying (2.8) and (2.11) (for suitable φ_F) and draw similar conclusions on T_{nJ} and T_{nB}^* . Besides $L_n(\cdot)$, other loss functions may as well be used in the multiparameter cases.

REFERENCES

- [1] Akahira, M. and Takeuchi, K. (1979). The concept of asymptotic efficiency and higher order asymptotic efficiency in statistical estimation theory. Lecture Note, Univ. Tokyo, Japan.
- [2] Athreya, K.B. (1987). Bootstrap of the mean i the infinite variance case. Ann. Statist. 15, 724-731.

- [3] Blyth, C.R. (1972). Some probability paradoxes in choice from among random alternatives. J. Amer. Statist. Assoc. 67, 366-373.
- [4] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Ann. Statist. 7, 1-26.
- [5] Efron, B. (1982). The Jackknife, Bootstrap and Other Resampling Methods, SIAM Publ., Philadelphia.
- [6] Ghosh, M. and Sen, P.K. (1989). Median Unbiasedness and Pitman Closeness. J. Amer. Statist. Assoc. 84, 1089-1091.
- [7] Hall, P. (1989) On efficient bootstrap simulation. Biometrika 76, 613-617.
- [8] Hall, P. (1989). Antithetic resampling for the bootstrap. Biometrika 76, 713-724.
- [9] Jureckova, J. (1984). M-, L- and R- estimators. In Handbook of Statistics, Vol. IV: Nonparametric Methods (eds. P.R. Krishnaiah and P.K. Sen). North Holland, Amsterdam.
- [10] Nayak, T. (1990). Estimation of location and scale parameters using generalized Pitman nearness criterion. J. Statist. Plan. Infer. 24, 259-268.
- [11] Pitman, E.J.G. (1937). The closest estimators of statistical parameters. Proc. Comb. Phil. Soc. 33, 212-222.
- [12] Sen, P.K. (1973) On fixed size confidence bands for the bundle strength of filaments. Ann. Statist. 1, 526-537.
- [13] Sen, P.K. (1977). Some invariance principles relating to jackknifing and their role in sequential analysis. Ann. Statist. 5, 316-329.
- [14] Sen, P.K. (1986). Whither jackknifing in Stein-rule estimation? Commun. Statist. Theor. Meth. 15, 2245-2266.

- [15] Sen, P.K. (1988). Functional jackknifing: Rationality and general asymptotics. Ann. Statist. 16, 450-469.
- [16] Sen, P.K. (1988). Functional approaches in resampling plans: A review of some recent developments. Sankhya, Ser. A 50, 394-435.
- [17] Sen, P.K. (1990). On the Pitman closeness of some sequential estimators. Sequen. Anal. 9.
- [18] Sen, P.K. (1991). Pitman closeness of statistical estimators: Latent years and the renaissance. In Issues in Statistical Inference: Essays in Honor of D. Basu (eds. M. Ghosh and P.K. Pathak), IMS Lecture Note Monogr. Ser. No. 17.
- [19] Sen, P.K., Kubokawa, T. and Saleh, A.K.M.E. (1989). The Stein paradox in the sense of the Pitman measure of closeness. Ann. Statist. 17, 1375-1386.