ASYMPTOTICS VIA SUB-SAMPLING BAHADUR-TYPE REPRESENTATIONS

by

Pranab Kumar Sen

Department of Biostatistics

University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1854

July 1988

# ASYMPTOTICS VIA SUB-SAMPLING BAHADUR-TYPE REPRESENTATIONS

By PRANAB KUMAR SEN

*University of North Carolina at Chapel Hill* .

SUMMARY. In the context of jackknifing, bootstrapping and other robust estimation procedures for general statistical parameters, one may encounter some statistics which may not be linear in the von Mises functional sense. Study of the asymptotic properties of these statistics by the conventional methods may require elaborate analysis and/or comparatively more stringent regularity conditions. In the context of robust estimation, often, the entire sample statistic is expressible in terms of similar sub-sample statistics. If the usual Bahadur-type representations hold for such sub-sample statistics, then the asymptotics for the whole sample statistic can be worked out neatly by some nearly linear analysis. This approach to the asymptotic theory is illustrated with several important examples.

## 1. INTRODUCTION

Based on a sample $X_1,\dots,X_n$ of size n, let $T_n = T(X_1,\dots,X_n)$ be a suitable statistic. In the asymptotic theory, one allows n to increase indefinitely with a view to simplifying the distributionl properties of $\{(T_n - a_n)/b_n\}$ , where $\{a_n\}$ and $\{b_n\}$ (nonnegative) are suitable sequences of real numbers. These results along with other refinements ( such as the weak as well as strong invariance principles ) are well covered in contemporary text boobs in large sample theory. But, in the majority of these texts, $T_n$ has been taken as a sum ( or average) of independent random variables (r.v.) or (forward or backward ) martingales or sub( or super-) martingales; the results continue to hold when $T_n$ can be approximated (in a suitable norm) by a version of a forward or backward (sub or super-) martingale . However, to make such approximations workable in the specific contexts, one may require either some

---

AMS (1980) *Subject Classifications*: 62E20, 62G99

*Key words and phrases*: *Asymptotic normality; Bahadur-type representations ; bootstrap methods; conditional quantile functions; empirical distributions; jackknifing; nonparametric regression; robust ( R- , M- and L-) estimators; statistical functionals.*

elaborate mathematical manipulations or some stringent regularity conditions.

Dealing with sample quantiles in large samples, Bahadur (1966) considered a remarkably general (and yet simple) representation, which for more than twenty years has played a key role in the asymptotic theory for general non-linear statistics. Viewed from a slightly general point, we may introduce the following notions on the *Bahadur representations* .

(i) **Weak Bahadur Representation (WBR)**: A sequence $\{T_n\}$ is said to have a weak Bahadur representation if there exists a score function $\phi$, such that

$$T_n - \theta = n^{-1} \Sigma_{i=1}^{n} \phi(X_i) + R_n , \tag{1.1}$$

where $\theta$ is a parameter ,

$$|R_n| = o_p(n^{-1/2}) , \text{ as } n \to \infty , \tag{1.2}$$

and $\phi(X)$ has mean 0 and a finite positive variance $\sigma_\phi^2$ . Thus, a possibly nonlinear statistics $T_n$ is expressible in terms of the linear component $n^{-1}\Sigma_{i=1}^{n} \phi(X_i)$ upto the order $o_p(n^{-1/2})$; the form of the score function $\phi(.)$ may depend on the underlying distribution function (d.f.) and/or other parameters associated with it.

For the specific case of sample quantiles, which are typically nonlinear functions of the sample observations, we may refer to Ghosh (1971) for a nice treatment of (1.1) and (1.2). For rank (R-) estimators of location/regression , robust ( M-) estimators of location/regression and linear functions of order statistics (i.e., L-estimators), this weaker form of Bahadur representation has been studied extensively in the literature; most of these developments are reported in Serfling (1980) and Sen (1981), among others.

(ii) **Strong Bahadur Representation (SBR)**: A sequence $\{T_n\}$ is said to have a strong Bahadur reprensentation if (1.1) holds and there exists a suitable sequence $\{\varepsilon_n\}$ of positive numbers, such that $\varepsilon_n \downarrow 0$ as $n \to \infty$ ( in a suitable manner) and

$$|R_n| = O( \varepsilon_n ) \text{ almost surely (a.s.), as } n \to \infty . \tag{1.3}$$

For sample quantiles, Bahadur (1966) showed that (1.1), (1.3) hold with $\varepsilon_n = n^{-3/4}\log n$ ; sharper results were obtained by Kiefer (1967) by some elaborate analysis. For von Mises' functionals, U-statistics and L-, M- and R-estimators, such SBR results are reported in Sen (1981).

(iii) **Bahadur Representation in the rth mean** (BRR): Suppose that (1.1) holds, and defining $\{\varepsilon_n\}$ as in (1.3), for some positive r,

$$E|R_n|^r = O(\varepsilon_n^{-r/2}), \quad \text{as } n \to \infty, \tag{1.4}$$

then $\{T_n\}$ is said to have a Bahadur representation in the rth $(r > 0)$ mean. For sample quantiles, Duttweiler (1973) considered such a BRR result for $r = 2$ and showed that $\varepsilon_n = O(n^{-3/4})$. This particular case of r=2 also arises in the well known projection results for nonlinear statistics [viz., Hoeffding (1948) and Hájek (1968)]. Let

$$S_n = \Sigma_{i=1}^n E\{T_n - \theta \mid X_i\} = \Sigma_{i=1}^n \phi_n(X_i), \tag{1.5}$$

where the form of the score function $\phi_n(.)$ may depend on the underlying d.f. or $\theta$. Then, we may write

$$T_n - \theta = S_n + R_n^*, \text{ say,} \tag{1.6}$$

which resembles (1.1)-(1.4). However, in (1.4), r may not be equal to 2, while in the projection result in (1.6), we have made use of $r = 2$. Moreover, in (1.6), the $\phi_n(X_i)$ may depend on n in a more involved manner (i.e., may not be equal to $n^{-1}\phi(X_i)$). Thus, the BRR may be more general than the usual projection result (with respect to the choice of $r > 0$), but may not be so with respect to the score function. As we shall see, the BRR plays the most important role in the asymptotic theory to be discussed here.

Note that the $\Sigma_{i=1}^n \phi(X_i)$ form a forward martingale sequence, while for identically distributed (i.d.) r.v.'s, the $n^{-1}\Sigma_{i=1}^n \phi(X_i)$ form a backward martingale sequence. Thus, the classical central limit theorems, weak or strong invariance principles, law of iterated logarithm and related results for martingales or reverse martingales can be conveniently incorporated in (1.1) along with the appropriate WBR, SBR or BRR result to yield parallel results for $\{T_n\}$. Our main objective is to explore this approach in a class of problems where the whole sample statistic $T_n$ can be expressed in terms of an average of sub-sample statistics of similar forms. In such a case, verifying (1.1) [and (1.2), (1.3) or (1.4)] for $T_n$ in a conventional manner may appear to be relatively harder, but for the sub-sample statistics these can be done relatively simply. The important role of the Bahadur-type representations in such a problem will be fully assessed here. Secondly, a sub-sampling scheme is also very appropriate in the context of nonparametric regression estimation problems with multivariate observations. Both the jackknifing and bootstrapping methods depend on the skillful choice of a subset of the observations and on resampling plans on this subset. Here also, the Bahadur-type representation plays a vital role, and we shall discuss this too.

## 2. ROBUSTIFICATION THROUGH A SUB-SAMPLING SCHEME

Suppose that $T_n$ is an estimator of $\theta$ , and it may not be linear in nature. If in the sample $(X_1,\ldots,X_n)$ there are a few outliers , their impact on the lack of robustness ( and efficiency ) of $T_n$ may depend appreciably on the functional form of $T_n$ ( viz., the sample mean or variance which attach quite large weights to the outliers ). In order to crub the influence of the outliers ( without changing the form of $T_n$ significantly ), it may be advisable to consider the following resampling scheme.

For a given $k : 1 \leq k \leq n$ , consider a subsample $X_{i_1},\ldots,X_{i_k}$ and compute $T_{(i_1,\ldots,i_k)} = T(X_{i_1},\ldots,X_{i_k})$. Note that there are $\binom{n}{k}$ possible subsets of $\{i_1,\ldots,i_k\}$ ( out of $\{1,\ldots,n\}$ ). Consider the model where in the sample there may be at most m outliers, such that m/n is small. Consider then the case where $k = k_n$ may depend on the sample size n, in such a way that $k_n$ and $n-k_n$ are both large ( and greater than the stipulated value of m ). Suppose that $J = \{j_1,\ldots,j_m\}$ stands for the index set for the outliers ( not known). If none of the $j_s$ is contained in the subset $\{i_1,\ldots,i_k\}$ , then $T_{(i_1,\ldots,i_k)}$ is least affected by the outliers. Similarly, if only one of the $j_s$ is contained in $\{i_1,\ldots,i_k\}$ , then $T_{(i_1,\ldots,i_k)}$ will be influenced by the outliers, but possibly to a smaller extent than in the case where multiple subscripts $(j_s)$ are present in $\{i_1,\ldots,i_k\}$ . In this way, we may classify the impact of the outliers on the $T_{(i_1,\ldots,i_k)}$ by the cardinality of the sets $J \wedge \{i_1,\ldots,i_k\}$ ( which may range from 0 to m if $k \geq m$). But neither m nor the set J is specified, and hence, we can not classify the subsets $\{i_1,\ldots,i_k\}$ in the light of the cardinality of $J \wedge \{i_1,\ldots,i_k\}$ . However, if we take an average over all possible $T_{(i_1,\ldots,i_k)}$ , the picture may become quite favorable. To see this point in a more visible context, we consider the following modefied estimator:

$$T^*_{n,k} = \binom{n}{k}^{-1} \Sigma_{\{1 \leq i_1 < \ldots < i_k \leq n\}} \quad T_{(i_1,\ldots,i_k)} , \tag{2.1}$$

where ( without any loss of generality ) we assume that for each $i_1,\ldots,i_k$ , $T_{(i_1,\ldots,i_k)}$ is a symmetric function of $X_{i_1},\ldots,X_{i_k}$ . Let $T^{(r)}_{n,k}$ stand for the average over all those $T_{(i_1,\ldots,i_k)}$ for which $J \wedge \{i_1,\ldots,i_k\}$ has the cardinality r, for r = 0,1,...,m. Then, we may rewrite (2.1) as

$$\begin{aligned} T^*_{n,k} &= \binom{n}{k}^{-1} \Sigma^m_{r=0} \binom{m}{r}\binom{n-m}{k-r} T^{(r)}_{n,k} \\ &= T^{(0)}_{n,k} + \binom{n}{k}^{-1} \Sigma^m_{r=1} \binom{m}{r}\binom{n-m}{k-r}[ T^{(r)}_{n,k} - T^{(0)}_{n,k} ] \\ &= T^{(0)}_{n,k} + \{n^{-1}mk\}\binom{n-1}{k-1}^{-1}\binom{n-m}{k-1}( T^{(1)}_{n,k} - T^{(0)}_{n,k}) + \{m(m-1)k(k-1)/2n(n-1)\}(T^{(2)}_{n,k} - \\ &\qquad T^{(0)}_{n,k}) + \ldots + \binom{n}{k}^{-1}\binom{n-m}{k-m}( T^{(m)}_{n,k} - T^{(0)}_{n,k} ) . \end{aligned} \tag{2.2}$$

Note that if k ($=k_n$ ) and n-k are both large compared to m , then in (2.2) the maximum weight is given to $T^{(0)}_{n,k}$ , and the weigts are decreasing ( and asymptotically negligible) with r (=1,...,m). Thus the robustness of the component $T^{(r)}_{n,k}$ has an important bearing on the robustness of $T^*_{n,k}$ . Keeping (2.2) in mind, we now assume that

$$\begin{aligned} &m \text{ is fixed and } k = k_n \text{ is such that } n-k_n \text{ and } k_n \text{ are both} \uparrow \text{ in } n \\ &\text{and } \quad n^{-1}mk_n \downarrow 0 , \text{ as } n \to \infty . \end{aligned} \tag{2.3}$$

Thus, if for the sub-sample statistics, we start with some robust estimators and then consider the modified estimator in (2.1), the robustness aspect remains in tact, and asymptotically, we would expect $T^*_{n,k_n}$ to behave as $T^{(0)}_{n,k_n}$ ( which is unaffected by the outliers). Our main interest lies in the asymptotic properties of the estimator $T^*_{n,k_n}$ in (2.1) for the outlier model for which (2.3) holds.

Note that for any (fixed) k ( $\geqslant 1$) , $T^*_{n,k}$ in (2.1) is a U-statistic [ c.f. Hoeffding (1948)] with kernel $T_{(1,\ldots,k)}$ of degree k. Hence, the asymptotic theory of $T^*_{n,k}$ can be studied directly with the aid of the extensive results on U-statistics available in the literature [ c.f. Serfling (1980), Sen (1981) and others]. The interesting point is that if $T_{(1,\ldots,k)}$ is itself a U-statistic corresponding to a kernel of degree d( $\geq$ 1) and if d does not depend on n ( also, d $\leq$ k ), then $T^*_{n,k}$ = $T_n$ , the U-statistic corresponding to the same kernel ( and sample size n ). This shows that for the entire class of U-statistics for which the kernels have degree not dependent on the sample size, the modified estimator in (2.1) agrees with the whole sample U-statistic ($T_n$) , and hence, no new methodology is needed to study the related asymptotic theory. However, this also brings out the deficiency of such U-statistics from the robustness point of view ( as may be verified with the aid of the sample mean and variance which are both U-statistics). From this robustness aspect, we may generally consider k= $k_n$ satisfying (2.3), and in such a case, the picture may be quite different. To apply the Hoeffding (1948) results on U-statistics when the degree (k) depends on n ( i.e., k = $k_n$ is increasing with n ), we may need additional uniformity conditions which may take away much of the charms of the elegant projection results in Hoeffding (1948). Similarly, a direct verification of the regularity conditions in Miller and Sen (1972) for the invariance principles for U-statistics ( and von Mises functionals ) when the degree $k_n$ increases with n may require extensive manipulations and/or more stringent regularity conditions. To illustrate the utility of the Bahadur-type representations, suppose now that (2.3) holds and for every $1 \leq i_1 < \ldots < i_k \leq n$ ,

$$T_{(i_1,\ldots,i_{k_n})} - \theta = k_n^{-1} \Sigma_{j=1}^{k_n} \phi(X_{i_j}) + R_{(i_1,\ldots,i_{k_n})} , \tag{2.4}$$

where, for some $r \geq 1$,

$$E|R_{(i_1,\ldots,i_{k_n})}|^r = o(n^{-r/2}) . \tag{2.5}$$

Then, by (2.1) and (2.4)-(2.5), we have

$$T^*_{n,k_n} - \theta = n^{-1}\Sigma_{i=1}^{n} \phi(X_i) + R^*_{n,k_n} \tag{2.6}$$

where

$$R^*_{n,k_n} = \binom{n}{k_n}^{-1} \Sigma_{\{1\leq i_1<\ldots<i_{k_n}\leq n\}} R_{(i_1,\ldots,i_{k_n})}$$

$$= E\{ R_{(1,\ldots,k_n)} \mid \mathcal{C}_n \} \tag{2.7}$$

and $\mathcal{C}_n = C(X_{n:1},\ldots,X_{n:n}; X_{n+j}, j \geq 1)$ denotes the tail sigma-field (and is nonincreasing in n); $X_{n:1},\ldots,X_{n:n}$ stand for the ordered r.v.'s corresponding to $X_1,\ldots,X_n$ , $n \geq 1$. Recall that

$$\begin{aligned} E|R^*_{n,k_n}|^r &= E\{ |E(R_{(1,\ldots,k_n)} \mid \mathcal{C}_n )|^r \} \\ &\leq E| R_{(1,\ldots,k_n)} |^r \\ &= o(n^{-r/2}), \text{ by (2.5)} . \end{aligned} \tag{2.8}$$

Hence,

$$R^*_{n,k_n} = o_p(n^{-1/2}), \tag{2.9}$$

so that by (2.6) and (2.9), we obtain that as $n \to \infty$,

$$T^*_{n,k_n} - \theta = n^{-1}\Sigma_{i=1}^n \phi(X_i) + o(n^{-1/2}) \text{ , in prob/rth mean.} \tag{2.10}$$

This first order Bahadur representation for $\{T^*_{n,k_n}\}$ can be incorporated in the study of the asymptotic properties of $\{T^*_{n,k_n}\}$ through the sequence $\{\phi(X_i);i \geq 1\}$ of i.i.d.r.v.'s. Thus, the crux of the problem is to verify (2.4)-(2.5). We shall illustrate this with some important examples ( having interest of their own) in the next section. In the rest of this section, we make some general comments on (1.1)-(1.4) with respect to their adaptability in (2.11).

First, in (2.4)-(2.5), we have restricted ourselves to $r \geq 1$ (although r may not be an integer or equal to 2). Note that for $r \in (0,1)$, $|x|^r$ is not a convex function, and hence, the first inequality in (2.8) may not hold. A proof of (2.9) would then require to show that the U-statistic on the right hand side of (2.7) is $o_p(n^{-1/2})$. Remember that for $r < 1$, this U-statistic may not be adaptable to a reverse martingale, and hence, the usual techniques may not work out here. Though this can be done by specific constructions in various specific cases, a general formulation may require extra regularity conditions.Secondly, in (2.10), the remainder term is $o(n^{-1/2})$ in probability as well as in the rth mean (for $r \geq 1$). The convergence in the rth mean may be useful in many problems requiring the uniform integrability of $n^{1/2}( T^*_{n,k_n} - \theta )$. Thirdly, we may note that the WBR in (1.1)-(1.2) for the sub-sample $T_{(i_1,\ldots,i_{k_n})}$ may not automatically ensure (2.10). The technical difficulty is due to the fact that there are $\binom{n}{k_n}$ possible remainder terms in (2.7); marginally each one may be $o_p(n^{-1/2})$ , but they are not all independent. Thus, we encounter a similar problem invonving a U-statistic where the kernel is $o_p(n^{-1/2})$, although it may not be integrable. Hence, to obtain (2.10) under (1.2), the basic problem is to study the weak convergence of a U-statistic ( without incorporating the $L_1$-norm) when the kernel is $o_p(n^{-1/2})$ and its degree may depend on n ( through $k_n$ ). Thus, with respect to the handling of the remainder term $R^*_{n,k_n}$ , the use of the WBR may generally require further scrutiny, and may not suffice without further regularity conditions. The SBR [originally considered by Bahadur (1966)] may be quite useful in this context. If we look at the basic proof of the Bahadur representation, we may observe that the a.s. order statement in (1.3) is actually based on a probability inequality statement with a rate of convergence either exponential in n, or, at least, $O(n^{-q})$, where $q(>1)$ is arbitrary. The use of such probability inequalities would generally lead us to (2.10) where $o_p(n^{-1/2})$ may be replaced by $o(n^{-1/2})$ a.s., as $n \to \infty$ . With this extra rate of convergence (implicit in the SBR) (1.2) may generally lead to (2.10), possibly in a stronger version [i.e., with $o(n^{-1/2})$ a.s. ]. It seems that the BRR in (1.1) and (1.4) is generally the most handy tool in verifying the weaker representation in (2.11).

## 3. SOME ILLUSTRATIVE EXAMPLES

First, consider a general class of robust estimators of location, treated in the finite sample setup in Sen (1964). For some nonnegative integer k, based on the sub-sample $X_{i_1},\ldots,X_{i_{2k+1}}$ ( of size 2k+1 ), let $\tilde{X}_{i_1,\ldots,i_{2k+1}}$ be the median, and let

$$T^*_{n,2k+1} = \binom{n}{2k+1}^{-1} \Sigma_{\{1\le i_1<\ldots<i_{2k+1}\le n\}} \tilde{X}_{i_1,\ldots,i_{2k+1}} . \tag{3.1}$$

Note that for k = 0, (3.1) reduces to the sample mean $\bar{X}_n$ $(= n^{-1}\Sigma_{i=1}^n X_i)$ (which is usually efficient but non-robust ), while for $k = [(n-1)/2]$, (3.1) reduces to a version of the whole sample median ( which is highly robust but may not be that efficient). Based on these considerations, Sen (1964) advocated the choice of k as a positive intger, and studied the relative efficiency of $T^*_{n,2k+1}$ for small values of k and n ($\le$ 10) when the underlying d.f. is taken to be normal. For a fixed k, (3.1) is actually a U-statistic for a kernel of degree 2k+1, and hence, its asymptotic theory follows readily from the general results in Hoeffding (1948), Miller and Sen (1972) and others. Apparently unaware of this work, Yanagawa (1969) considered the same estimator and studied its robustness properties. Some closely related estimators (of a general quantile ) are due to Kaigh and Lachenbruch (1982) and Harrell and Davis (1982). All these versions relate to a fixed k, and they are very close to each other and the classical sample quantile estimator : A general result to this effect is due to Yoshizawa, Sen and Davis (1985). In this context, note that if $X_{n:1} \leq \cdots \leq X_{n:n}$ stand for the ordered r.v.'s corresponding to $X_1, \ldots, X_n$, then for every k: $0 \leq k \leq (n-1)/2$ ,

$$T^*_{n,2k+1} = \binom{n}{2k+1}^{-1} \Sigma_{i=k+1}^{n-k} \binom{i-1}{k}\binom{n-i}{k} X_{n:i} , \tag{3.2}$$

so that (3.1) is in fact an L-estimator too; for $k \geq 1$, (3.2) attaches smaller weights to the extreme order statistics and larger weights to the central ones, and hence, behaves more robustly than the sample mean (which corresponds to $k = 0$ ).

Consider now the motivation of Section 2 and allow $k = k_n$ to depend on n in such a way that (2.3) holds. This will allow us to study the asymptotic properties in the presence of outliers. We appeal to the basic results of Bahadur (1966) along with the parallel quadratic mean convergence version [ c.f. Duttweiler (1973)] and obtain that the BRR result in (1.1) and (1.4) holds with $\phi(X_i) = I(X_i \geq \theta) - 1 + F(\theta)$, whenever

$$k_n \sim n^{\lambda} , \text{ for some } \lambda > 2/3 . \tag{3.3}$$

Actually, we may even let $k_n \sim n^{2/3}(\log n)^2$ . As such the representation in (2.10) can readily be obtained under (3.3). In this specific example, although for $k = k_n$ ( increasing with n ) , Hoeffding's (1948) U-statistics theory may require some extra manipulations ( for handling the projection result ), viewed from the L-estimation point of view, for any such $k_n$ , (3.2) represents an L-estimator with smooth weights, and hence, the asymptotic theory may also be studied with the aid of the general results discussed in Sen (1981, Ch. 7) and Serfling (1980), among others. The main advantage of using the BRR result and (2.10) is having a representation in terms of a set of i.i.d.r.v.'s (plus a remainder term) in a most simple manner.

Let us next consider a modification of the classical L-, M- and R-estimators of location ( to induce robustness in the same sense as in Section 2 ). The case with the L-estimators is most apparent. For any subsample size $k_n$ ( $\leq n$), an L-estimator is a linear combination of functions of its order statistics where the weights depend only on $k_n$ . Thus, if we define $T^*_{n,k_n}$ as in (2.1) with the individual $T_{(i_1,\ldots,i_{k_n})}$ as L-estimators (based on the common scores), then $T^*_{n,k_n}$ remains as an L-estimator with adjusted weights depending on $(n,k_n)$. As such, its asymptotic theory can again be studied with the help of the general results provided in Serfling (1980) and Sen (1981), among others. The situation is different with M- and R-estimators.

By very construction ( i.e., the choice of appropriate score functions ), the M- and R-estimators are robust (against outliers ) and efficient too. Hence, there may not be a profound need to choose $k = k_n$ small compared to n . Keeping this in mind, we consider a modified estimator as in (2.1) with $k = k_n = n-1$ ( where the $T_{(i_1,\ldots,i_{n-1})}$ are the M- or R-estimators based on the sample $(X_{i_1},\ldots,X_{i_{n-1}})$ of size n-1. That is, we consider an <u>one-step iterated</u> estimator

$$\hat{\theta}^*_n = n^{-1} \Sigma^n_{i=1} \hat{\theta}_{\underset{\sim}{n}\setminus i} , \tag{3.4}$$

where $\underset{\sim}{n} = \{1,\ldots,n\}$ and $\underset{\sim}{n}\setminus i = \{1,\ldots,i-1,i+1,\ldots,n\}$, for $i = 1,\ldots,n$. [Instead of the one-step version, we could have constructed an m-step version for any (fixed) $m \geq 1$, but the conclusions to follow would remain the same. ] Note that for the complete sample (M- or R- ) estimator $\hat{\theta}_n$ , there exists a score function $\phi(x)$, $x \in E$ , such that (1.1)-(1.4) hold ; we may refer to Chapter 8 ( and 10) of Sen (1981) where the weak ( and strong) representations have been systematically discussed. As such, for the individual $\hat{\theta}_{\underset{\sim}{n}\setminus i}$ , we have no problem in claiming that (1.1)-(1.4) hold under the same regularity conditions. As such, we obtain that under the regularity conditions studied in detail in Chapter 10 of Sen (1981), for each i ( =1,...,n), as n increases,

$$\hat{\theta}_{\underset{\sim}{n}\setminus i} - \theta = (n-1)^{-1} \Sigma_{j \in \underset{\sim}{n}\setminus i} \phi(X_j) + R_{n,i} \tag{3.5}$$

where the $\phi(X_j)$ have zero mean and a finite variance, and the $R_{n,i}$ are $o(n^{-1/2})$ in the rth mean ( $r > 0$, arbitrary) as well a.s. as $n \to \infty$ ( and hence, in probability too). Going down (2.4) through (2.9), we may therefore conclude that (2.10) holds for $\hat{\theta}_n^*$, where $o_p(n^{-1/2})$ may as well be replaced by $o(n^{-1/2})$ a.s. as $n \to \infty$ or by $o(n^{-1/2})$ in the rth mean for any (fixed) $r > 0$. On the other hand, a direct approach to verify (2.10) for $\hat{\theta}_n^*$ would have been quite involved.

We may remark that if in (2.1) each $T_{(i_1,\ldots,i_{k_n})}$ is a U-statistic [ based on $X_{i_1},\ldots,X_{i_{k_n}}$ ] corresponding to a kernel of degree m (fixed), then $T^*_{n,k_n}$ is the complete sample U-statistic (based on the same kernel); in this case, the asymptotic distribution theory of $T^*_{n,k_n}$ does not depend on $k_n$ (but on m ), and the classical results in Hoeffding (1948) apply directly. Nevertheless, the decomposition in (2.2) gives a clear picture on the robustness of U-statistics in the presence of some outliers. For von Mises' functionals, the results, albeit close, are not exactly the same [ as the individual $T_{(i_1,\ldots,i_{k_n})}$ may involve more than one kernel with the maximum weight given to a particular one]. In such a case, the proposed modification may work out better. In the same vein, we may consider a general statistical functional $\theta = T(F)$ where F is the d.f. of the $X_i$ ( assumed to be i.i.d.r.v.'s). A natural estimator of T(F) is the sample counterpart $T_n = T(F_n)$, where $F_n$ is the sample (empirical) d.f. of $X_1,\ldots,X_n$. If T(G) is first order Hadamard (compact) differentiable (at F) and if $T_1(F;x)$ stands for the influence function of T(.) at F, then [ viz., Sen (1988)] we have

$$T_n - T(F) = n^{-1}\Sigma_{i=1}^n T_1(F;X_i) + o(\ ||F_n - F||\ ), \tag{3.6}$$

where $||.||$ stands for the sup-norm, and from the classical results on the Kolmogorov-Smirnov statistics, it is known that $||F_n - F|| = O_p(n^{-1/2})$. In fact, it is also known that $||F_n - F|| = O(n^{-1/2})$ in the rth mean for any (fixed) $r > 0$. As such, by choosing an $r \geq 1$ and appealing to the results in Section 2, we immediately conclude that for the $T^*_{n,k_n}$ ( corresponding to $T_n$) , (2.10) holds in the rth mean as well. Thus, as in (2.1)-(2.2), a robust version of $T_n$ may be constructed and its asymptotic theory can readily be studied with the aid of the results in Section 2. In this context, however, we choose $k_n = O(n)$ [ as in (3.4)]. If $k_n = o(n)$ [ as in (3.3)], then the first order decomposition in (3.6) may not suffice. If we assume that T(G) is second order Hadamard differentiable at F, then proceeding as in Sen (1988), we have parallel to (3.6)

$$T_n - T(F) = n^{-1}\Sigma_{i=1}^{n} T_1(F;X_i) + U_n^{(2)} + o(\ ||F_n - F||^2\ ), \tag{3.7}$$

where $U_n^{(2)}$ is a U-statistic (of degree 2 ) and is stationary of order 1 (i.e., var($U_n^{(2)}$) is $O(n^{-2})$). In this case, if we assume that $k_n$ is so chosen that $n^{-1}k_n^2$ goes to $\infty$ as $n \to \infty$ [ i.e., (3.3) holds for some $\lambda > 1/2$ ], then (2.10) holds in the rth mean ( for r = 2 ). It may be remarked that both M- and R-estimators may be characterized as $T(F_n)$ , although for such a functional in order to justify (3.6) [or (3.7)] we may need some extra regularity conditions which may not be necessary with the direct approach considered in Sen (1981) or more generally in Jurečková (1985). For L-estimators and M-estimators with bounded (and smooth ) score functions these extra regularity conditions hold, while for R-estimators, such a bounded score function may exclude the case of Normal scores ( or log rank ) estimators, and hence, we would rather advocate the direct approach.

## 4. SUB-SAMPLING SCHEMES IN NONPARAMETRIC REGRESSION ESTIMATION

Consider the following simple regression model in a purely nonparametric setup. Let $(X_i,Y_i)$, $i \geq 1$ be a sequence of i.i.d.r.v.'s with a bivariate d.f. $\Pi(x,y)$ , defined on $E^2$. Let $F(x) = \Pi(x,\infty)$ be the marginal d.f. of X and let $G(y|x)$ be the conditional d.f. of Y given X = x. Consider a suitable functional

$$\theta_x = T_x = T(G(.\ |x)) \text{ of the conditional d.f. } G(.\ |x). \tag{4.1}$$

Notable forms of $T_x$ are the (conditional) mean of Y given X = x ( i.e., the mean of the d.f. $G(.|x)$), the median ( or some other quantile ) of $G(.|x)$ and other measures of location of this conditional d.f. $G(.|x)$. The basic problem is to estimate $\theta_x$ in a nonparametric setup (i.e., without assuming any structural form for $\theta_x$ ). In this model, the $X_i$ can as well be vector valued. But, for simplicity of presentation, we take them to be real valued r.v.'s.

Suppose now that we want to estimate $\theta_x$ at a given point $x_0$ . The problem can be handled very conveniently when $x_0$ lies well within the range spanned by the $X_i$ , i.e., $0 < F(x_0) = p_0 < 1$. For a sample $X_1,\ldots,X_n$ ( of the $X_i$ ) of size n, we let

$Z_i = |X_i - x_0|$, for $i = 1,\ldots,n$, and let

$0 \le Z_1^* \le \ldots \le Z_n^*$ be the associated order statistics. (4.2)

Let $\{k_n\}$ be a sequence of positive integers such that

$k_n \to \infty$ but $n^{-1}k_n \downarrow 0$ as $n \to \infty$ . (4.3)

Corresponding to the order statistics $Z_1^*,\ldots,Z_{k_n}^*$ , the induced order statistics ( for the $Y_i$ ) are denoted by $Y_1^o,\ldots,Y_{k_n}^o$ , respectively. [ Note that although the $Z_i^*$ are ordered, the $Y_j^o$ need not be so.] Let then

$G_n^*(y) = k_n^{-1} \Sigma_{j \le k_n} \; I(Y_j^o \le y)$, $y \in E$ , (4.4)

be the empirical d.f. based on the subsample $(Y_1^o,\ldots,Y_{k_n}^o)$. Note that this subsample is not selected at random; rather, its constrcution is based on the ordering of the $Z_i$ in (4.2). Then, a natural estimator of $\theta_{x_0}$ is

$T(G_n^*) = \hat{\theta}_n(x_0)$ , say. (4.5)

Note that $x_0$ is also implicit in $G_n^*(.)$ through the construction of the $Z_i^*$ . In this context, it may be noted that given the $Z_i^*$ , the $Y_i^o$ are conditionally independent ( but not necessarily i.d.), and under (4.3), $G_n^*$ consistently estimates $G(.|x_0)$, although at a rate slower than $n^{-1/2}$ . Thus, if the functional $T(G)$ is sufficiently smooth ( in the sense of (3.6) or (3.7)), then the plausible weak convergence of $k_n^{1/2}( G_n^* - G(.|x_0) )$ may be incorporated in the study of the asymptotic properties of $\hat{\theta}_n(x_0)$. However, such a smoothness condition may not be taken for granted for all functionals ( for example, the sample quantiles ). Also, there is a need to choose $k_n$ in such a way that one has an asymptotically optimal choice (within a given class of functionals). For these reasons, both jackknifing and bootstrap methods have been proposed by various workers. For a conditional quantile ( i.e., a quantile of $G(.|x_0)$), Bhattacharya and Gangopadhyay (1988) have shown that the asymptotic theory works out neatly when $k_n \sim n^\lambda$ , for some $\lambda \le 4/5$ ; for $\lambda = 4/5$, there may be a comparable (unknown) bias term in the asymptotic normality result, and hence, we may have to limit $\lambda$ to $< 4/5$ [see Gangopadhyay and Sen (1988)]. In the current study, we would like to bring the relevance of the SBR and BRR for the study of the asymptotic theory relating to this especial subsampling scheme.

Note that in Sections 2 and 3, we have considered a subsampling scheme which is essentially related to the simple random sampling without replacement (SRSWOR) scheme. The situation is different here. For estimating the unknown $G(.|x_o)$, we have considered $G_n^*$ in (4.4), where the subset $\{Y_1^o,\ldots,Y_{k_n}^o\}$ of the $Y_i$ is chosen on the 'nearest neighborhood' principle for the $X_i$ ( relative to the point $x_o$ ). Although given the $Z_i^*$, the $Y_j^o$ are conditionally independent ( but not necessarily i.d.), they are unconditionally not independent. Secondly, the $Z_i^*$ being the order statistics ( corresponding to the $Z_i$ )are not independent too, and we have deliberately chosen the lower extremity. These two technical points raise issues regarding the adaptability of the standard asymptotic methods in this nonparametric regression problem. Bahadur-type representation results play a very important role in this problem.

Recall that for the classical empirical d.f. $F_n$ ( corresponding to the true d.f. F) defined on E, whenever the density function (f) is positive at a quantile $\xi$ ( where $F(\xi) = p : 0 < p < 1$), and f' is bounded, the classical Bahadur (1966) representation asserts that as $n \to \infty$ ,

$$\sup\{ |F_n(x)-F(x) - F_n(\xi) + F(\xi)| : |x - \xi| \le (n^{-1}\log n)^{1/2} \} = O(n^{-3/4}\log n) \text{ a.s.} \quad (4.6)$$

Also, the sample p-quantile $\tilde{X}_n$ lies in $[\xi - (n^{-1}\log n)^{1/2}, \xi + (n^{-1}\log n)^{1/2}]$ a.s., as $n \to \infty$ , and hence, we have

$$\tilde{X}_n - \xi + (nf(\xi))^{-1} \Sigma_{i=1}^n [ I(X_i \le \xi) - p ] = O(n^{-3/4}\log n) \text{ a.s., as } n \to \infty . \quad (4.7)$$

Although the picture is more complicated in the current situation, (4.6) and (4.7) extend to $G_n^*$ in a very natural way ( where the role of n has to be replaced by $k_n$ ). This extended Bahadur representation for the nearest neighborhood sub-sampling scheme enables us to express

$$\hat{\theta}_n(x_o) - \theta_{x_o} = k_n^{-1} \Sigma_{j \le k_n} \phi(U_{ni}) + o(k_n^{-1/2}) \text{ a.s., as } n \to \infty , \quad (4.8)$$

where the $U_{ni}$ are conditionally i.i.d.r.v.'s and the score function $\phi(U_{ni})$ has zero ( conditional) mean. This conditional Bahadur-representation yields the desired asymptotic theory in a conditional setup, and the use of the Hewett-Savage zero-one law along with other standard tools in asymptotic theory provide the passage to the unconditional setup. For details, we may refer to Gangopadhyay and Sen (1988). In passing, we may also remark that in this conditional quantile or nonparametric regression problem, the choice of $k_n$ is essntially tied down to the smoothness conditions on F(.), G(y x) as well as their derivatives, and the use of the Bahadur representation provides a clear picture on the so called optimal rate for $k_n$ ($\sim n^{\lambda}$, for $\lambda < 4/5$). There is a subtle difference between the situations encountered in Section 2 and here. In (2.3), we may ideally take $k_n \sim \varepsilon n$, for some small $\varepsilon$ ( > 0), and this was illustrated in Section 3 with the choice of $k_n = n - 1$ . However, in the current context, a choice of $k_n = O(n)$ [ or $\sim \varepsilon n$ for some small $\varepsilon$ (> 0) ] may not lead to the general asymptotics. In Sections 2 and 3, the normalizing functions depend on n, while here it depends on $k_n$ , and therefore, we may have a much slower rate here. A delicate balance between the bias ( larger for higher order choice of $k_n$ ) and the rate of convergence ( faster for higher order choice of $k_n$ ) has to be enforced, and in that sense, $k_n \sim n^{\lambda}$ , $\lambda < 4/5$, seems to be a feasible solution. We may not be able to take $k_n \geq n^{\lambda}$ with $\lambda \geq 4/5$.

Nonparametric regression models have also been framed along the lines of the classical parametric linear models wherein the normality of the distribution of the error components have been replaced by an arbitrary ( continuous) distribution. In this context, the classical linear rank statistics may be used to yield robust (R-) estimators of regression parameters. Similarly, robust M-estimators of regression parameters have also been considered by a host of workers. We may refer to Puri and Sen (1985) and Jurečková (1985), for some of these details. In this context too, the Bahadur representation in (4.6), as extended to non-i.d. r,v.'s [ see Sen and Ghosh (1972)] plays a fundamental role. For such robust R- and M-estimators of regression parameters, if we desire to induce more robustness, we may consider as in Section 3, an one-step iterated version. For such versions, the asymptotic theory would follow along the lines sketched in Section 3.

# REFERENCES


BAHADUR, R.R. (1966). A note on quantiles in large samples. Ann. Math. Statist.37, 577-580.

BHATTACHARYA, P.K. and GANGOPADHYAY, A.K. (1988). Kernel and nearest neighborhood estimation of a conditional quantile. Univ. Calif.,Davis, Div. Statist. Tech. Rep. No. 104 ( to appear in the Ann. Statist.)

DUTTWEILER, D.L. (1973). The mean square error of Bahadur's order statistic approximation. Ann. Statist. 1, 446-453.

GANGOPADHYAY, A.K. and SEN,P.K. (1988). Bootstrap confidence intervals for conditional quantile functions. Univ.North Carolina, Inst. Statist. Mimeo Rep. No. 1765.

GHOSH, J.K. (1971). A new proof of the Bahadur representation of quantiles and an application. Ann. Math. Statist. 42, 1957-1961.

HÁJEK, J. (1968). Asymptotic normality of simple linear rank statistic under alternatives. Ann. Math. Statist. 39, 325-347.

HARRELL, F.E. and DAVIS, C.E. (1982). A new distribution-free quantile estimator. Biometrika 69, 635-640.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19, 293-325.

JUREČKOVÁ, J. (1985). Representation of M-estimators with the second order asymptotic distribution. Statist. Decs. 3, 263-276.

KAIGH, W.D. and LACHENBRUCH, P.A. (1982). A generalized quantile estimator. Commun. Statist. Theor. Meth.A 11, 2217- 2238.

KIEFER, J. (1967). On Bahadur's representation of sample quantiles. Ann. Math.Statist. 38, 1323-1342.

MILLER, R.G.JR. and SEN, P.K. (1972). Weak convergence of U-statistics and von Mises' differentiable statistical functions. Ann.Math. Statist. 43, 31-41.

PURI, M.L. and SEN, P.K. (1985). Nonparametric Methods in General Linear Models. John Wiley, New York.

SEN, P.K. (1964). On some properties of rank weighted means. Jour. Indian Soc. Agri. Statist. 16, 51-61.

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

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

SEN,P.K. and GHOSH, M. (1972). On strong convergence of regression rank statistics. Sankhya, Ser A 34, 335-348.

SERFLING, R.J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley, New York.

YANAGAWA, T. (1969). A small sample robust competitor of Hodges-Lehmann estimator. Bull. Math. Statist. 13, 1-14.

YOSHIZAWA, C.N., SEN, P.K. and DAVIS, C.E. (1985). Asymptotic equivalence of the Harrell-Davis median estimator and the sample median. Commun. Statist. Theor. Meth. A14, 2129-2136.