FUNCTIONAL APPROACHES IN RESAMPLING PLANS: A REVIEW OF SOME RECENT DEVELOPMENTS

Pranab Kumar Sen

#1771

1988
FUNCTIONAL APPROACHES IN RESAMPLING PLANS: A REVIEW OF SOME RECENT DEVELOPMENTS

By PRANAB KUMAR SEN

University of North Carolina at Chapel Hill

SUMMARY. The jackknife, bootstrap and some other resampling plans are commonly used to estimate (and reduce) the bias and sampling error of statistical estimators. In these contexts, general (differentiable) statistical functionals crop up in a variety of models (especially, in nonparametric setups). A treatise of statistical functionals in resampling plans is considered along with a review of some of these recent developments.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be $n$ independent and identically distributed random variables (i.i.d. r.v.) with a distribution function (d.f.) $F$, and let $T_n = T(X_1, \ldots, X_n)$ be a suitable statistic. For example, $T_n$ may be a suitable estimator of a parameter $\theta = \theta(F)$ which is a functional of the d.f. $F$, or $T_n$ may be a test statistic for testing a plausible null hypothesis concerning such a parameter $\theta(F)$. In the classical theory of statistical estimation, the distribution theory of such estimators occupies a central position. All the conventional criteria for studying the optimality properties of estimators depend on their sampling distributions or at least on some related measures. In a parametric setup, one assumes that $F$ has a given functional form involving some unknown parameters, so that $\theta(F)$ may as well be expressed in terms of these parametric quantities. In a variety of situations, for a given form of $F$, it may be possible to derive the exact distribution of an estimator, although such a task may become prohibitively laborious if $F$ does not belong to the exponential family and/or $n$ is large. In a nonparametric setup, the form of $F$ is unspecified, and hence, the distribution of $T_n$ is generally unknown. An exception is the case of some nonparametric test statistics which are distribution-free under suitable hypotheses of invariance where mostly for


Keywords and phrases: Asymptotic normality; bias; bootstrap; cross validation; delete-d jackknifing; functional jackknifing; generalized functionals; Hadamard differentiability; half-sample methods; influence function; jackknifing; L-,M- and R-estimators; nonparametric regression; outliers; pseudovariables; robustness; statistical functionals; Stein-rule estimation; variance estimation.

Short Title: FUNCTIONAL APPROACHES IN RESAMPLING PLANS.
small values of $n$ the exact distributions may be obtained by enumerations, although
the task becomes unmanageable for large values of $n$. There is thus a genuine need to
estimate the various characteristics of sampling distributions of competing estimators
or statistics; among these, the bias and mean square error criteria are most notables.

Resampling plans are generally designed to estimate the bias and mean square error
of suitable estimators, although they may as well be used to estimate their sampling
distributions in a variety of cases. Among these resampling plans, jackknifing was
introduced by Quenouille (1956) primarily for the reduction of bias of a (non-linear)
estimator. This method of resampling generates pseudovariables which also provide a
strongly consistent estimator of the mean square error of the estimator (and its
jackknifed version). Often, a statistic $T_n$ may be expressed as a functional $T(F_n)$ of
the sample (empirical) d.f. $F_n$. For smooth functionals, $T(F_n)$ can be expressed in
terms of a linear one and a residual term converging to zero at a certain rate with $n$, so
that the asymptotic bias and mean square error of $T(F_n)$ can be assessed in a rela-
tively simple manner. Both the infinitesimal jackknife and delta methods are based on
this decomposition, and it is not surprising [viz., Efron (1982)] that these two methods
yield identical results on the estimated mean square errors. A clever idea in resampling
plans, due to Efron (1979), is the popular bootstrap method. As we shall see later on,
there are certain situations where a bootstrap method may workout better than the
others, although there are other situations where jackknifing may be more appealing.
Efron (1982) contains a very comprehensive account of some of these resampling plans.

Albeit the emphasis on bias reduction and estimation of mean square error in a
resampling plan, there is a greater need to have a closer look from the point of view
of robustness. For instance, in the jackknife method, to filter robustness under jack-
knifing, one should start with a robust $T_n$; otherwise, the pseudovariables generated
by jackknifing being more vulnerable to outliers or error contaminations may lead to
a less robust jackknifed version. A similar situation arises with the bootstrap method.
Robustness aspects of the sample d.f. $F_n$ dominates the scenario in this case.

The past twenty-five years have witnessed a phenomenal growth of the literature on
(differentiable) statistical functionals and their impact on robust statistical inference. In the context of resampling plans, such functionals have been considered by Hinkley and Wang (1980), Babu and Singh (1984), Parr (1985) and Sen (1988a), among others. Resampling plans have also been adopted for nonparametric regression models; we may refer to Bhattacharya and Gangopadhyay (1988) where other references are cited. In the current study, we like to lay proper emphasis on the robustness aspects of statistical functionals in the context of various resampling plans. The preliminary notions on these resampling plans are introduced in Section 2; statistical functionals are formulated in Section 3. Section 4 contains a critical appraisal of statistical functionals in various resampling plans. Generalized functionals are treated in Section 5. Functionals for the nonparametric regression model are considered in Section 6. The concluding section deals with some general remarks and open problems.

2. RESAMPLING PLANS: RATIONALITY AND PRELIMINARY NOTIONS

We start with the observation that \( T_n = T(\mathbf{X}_1, \ldots, \mathbf{X}_n) \) may not be an unbiased estimator of \( \theta = \theta(F) \), and it may be possible to write

\[
E(T_n) = \theta(F) + n^{-1}a_1(F) + n^{-2}a_2(F) + \cdots \tag{2.1}
\]

where the \( a_j(F) \) are (unknown) functionals of the d.f. \( F \) (this is similar to what Quenouille (1956) had proposed earlier). Suppose that our primary goal is to reduce the bias without changing the nature of \( T_n \) to any substantial degree. Let us denote by

\[
T^{(i)}_{n-1} = T(X_{i}, X_{i+1}, \ldots, X_n), \quad i = 1, \ldots, n. \tag{2.2}
\]

Then, by (2.1) and (2.2), we have

\[
E(T^{(i)}_{n-1}) = \theta(F) + (n-1)^{-1}a_1(F) + (n-1)^{-2}a_2(F) + \cdots, \tag{2.3}
\]

so that

\[
E(nT_n - (n-1)T^{(i)}_{n-1}) = \theta(F) - a_2(F)/n(n-1) + O(n^{-3}), \quad i = 1, \ldots, n. \tag{2.4}
\]

With this observation, we may define the pseudovariables by

\[
T_{n,i} = nT_n - (n-1)T^{(i)}_{n-1}, \quad \text{for } i = 1, \ldots, n, \tag{2.5}
\]

and the jackknifed estimator

\[
T_n^* = n^{-1} \sum_{i=1}^n T_{n,i} = T_n + (n-1)n^{-1} \sum_{i=1}^n (T_n - T^{(i)}_{n-1}). \tag{2.6}
\]

Clearly, by (2.4) and (2.6),

\[
E(T_n^*) = \theta(F) - a_2(F)/n(n-1) + O(n^{-3}), \tag{2.7}
\]
so that by having the adjustment for bias (i.e., the second term on the right hand side
of (2.6)), the jackknifed estimator has bias of the order \( n^{-2} \), compared to \( n^{-1} \) in
(2.1). The resampling scheme in this context is the simple random sampling without
replacement (SRSWOR) of size \( n-1 \) from the basic sample of size \( n \). One may also consider
a SRSWOR scheme of size \( n-d \) out of \( n \), for some \( d \geq 1 \) (known as the delete \( d \) jackknifing
scheme) and obtain parallel estimates. The pseudovariables in (2.5) have been
incorporated in the formulation of the Tukey estimator of the variance:

\[
V_n^* = (n-1)^{-1} \sum_{i=1}^{n} (T_{n,i}^* - \bar{T}_n)^2,
\]

and under appropriate regularity conditions, it has been claimed that

\[
n^{3/2}(\bar{T}_n^* - \theta(F))/(V_n^*)^{3/2} \overset{D}{\to} N(0, 1), \text{ as } n \text{ increases.}
\]

(2.9)

The Tukey conjecture is a step further: The normal law in (2.9) can be replaced by
the Student t-distribution with \( n-1 \) degrees of freedom (DF) for moderately large
values of \( n \). But, the conjecture is not generally true and can be easily verified
by various examples (relating to nonlinear statistics). It may also be remarked that
Quenouille's motivation for using the specific form in (2.5) and (2.6) was geared by
(2.1) and (2.3), while some other workers [including Efron (1982, p. 10)] have rather
vaguely argued against it. However, there is another simple interpretation for this
jackknifing [viz., Sen (1977)] which is presented below.

For simplicity, let us assume that the \( X_i \) are real valued, and for every \( n(\geq 1) \),
let \( X_1, \ldots, X_n \) be the order statistics corresponding to \( X_1, \ldots, X_n \). Let \( C_n =
C(X_1, \ldots, X_n; X_{n+j}, j \geq 1) \) be the sigma-field generated by \( (X_1, \ldots, X_n) \)
and the \( X_{n+j}, j \geq 1 \), so that \( C_n \) is nonincreasing in \( n \). Note that given \( C_n \), \( X_1, \ldots, X_n \) are
interchangeable r.v.'s and they assume all possible permutations of \( X_1, \ldots, X_n \)
with the equal (conditional) probability \( (n!)^{-1} \). This is the natural justification
for adapting the SRSWOR plan leading to the Quenouille jackknifing procedure, and,
in fact, we have [as in Sen (1977)]

\[
T_n^* = T_n + (n-1)E\{T_n - T_{n-1} \mid C_n\} \text{ almost everywhere (a.e.)} \quad (2.10)
\]

Thus, if \( \{T_n, C_n ; n \geq n_0\} \) forms a reversed martingale sequence, then \( T_n^* = T_n \). Otherwise,
Quenouille's jackknifing consists in adding on the (bias) correction factor

\[
T_n^* - T_n = (n-1)E\{T_n - T_{n-1} \mid C_n\}, \quad n \geq n_0.
\]

(2.11)
so that the SRSWOR plan is a natural one for the formulation of the Tukey estimator of
the variance in (2.8). We may remark further that for the entire class of unbiased
estimators, jackknifing has no role in the bias reduction, although the Tukey variance
estimator may have the strong consistency property [ Sen(1977)] which alone is of
sufficient interest in various statistical inference problems. This simple interpreta-
tion in (2.11)-(2.12) is, however, limited to the case of exchangeable r.v.'s. In the
so called unbalanced designs or for non-i.d. observations, this may not workout [ viz.,
Wu(1986) and Shao(1988)]. The basic results in (2.7) and (2.9) extend directly to the
multivariate case (i.e., when \( T_n \) is itself a p-vector, for some \( p \geq 1 \)), and with an
extended definition of \( C_n \), the interpretation in (2.11)-(2.12) remains in tact.

Coming back to (2.9), we need to probe the following three results:

(i) Asymptotic normality of \( T_n \) : There exists a positive finite (and possibly
unknown ) constant \( \sigma \), such that as \( n \) increases,

\[
n_{\frac{1}{2}}( T_n - \Theta(F) )/\sigma \xrightarrow{d} N(0, 1);
\]

(ii) Consistency of \( V_n^* \) : As \( n \) increases,

\[
V_n^*/\sigma^2 \rightarrow 1 \quad \text{in probability.}
\]

(iii) Asymptotic equivalence of \( T_n \) and \( T_n^* \) : \( T_n \) and \( T_n^* \) are asymptotically equivalent
upto the order \( o(n^{-\frac{1}{2}}) \), i.e., as \( n \) increases,

\[
n_{\frac{1}{2}} | T_n - T_n^* | \rightarrow 0 \quad \text{in probability.}
\]

[ Recall that (2.13), (2.14), (2.14) and the Slutsky theorem ensure (2.9).] In the
literature, results stronger than (2.13) through (2.15) have been established under
diverse regularity conditions. For example, the asymptotic normality result has been
strengthened to weak or strong invariance principles (leading to Wiener processes
approximations), the weak consistency result to an almost sure convergence result,
and finally, (2.15) to a much stronger form that

\[
n( T_n - T_n^* ) = O(1) \quad \text{almost surely (a.s.), as } n \rightarrow \infty.
\]

We may refer to Sen(1977,1988b) and Parr(1983,1985), among others, for some discussions
on these stronger results. We shall discuss some of these in later sections. The
asymptotic normality result in (2.13) is of quite general form and it holds under
fairly general regularity conditions. The consistency result in (2.14) may require
additional regularity conditions, while (2.15) may demand even more stringent conditions. These will be discussed in a later section, the following example is noteworthy.

**Example 1.** Let \( X_1, \ldots, X_n \) be i.i.d.r.v. with \( N(0, \tau^2) \) distribution. Let \( T_n \) be the sample mean. Then it is easy to verify that \( T_n^* = T_n \) and \( V_n^* = s_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - T_n)^2 \). In this case (2.13), (2.14) and (2.15) all hold. Next, we take \( T_n = \text{med}(X_1, \ldots, X_n) \). Then (2.13) holds with \( \sigma^2 = \tau \pi/2 \). However, in this case (2.14) does not hold. If, on the other hand, we adopt a delete-d jackknifing procedure, then a version of \( V_n^* \) can be obtained such that (2.14) and (2.15) both hold when \( d \) is not small [viz., Wu(1986)].

Partially inspired by this example, we may set an outline of the delete-d jackknife method and the related resampling plan as follows. From the basic sample of size \( n \) (> \( d \)), by SRSWOR, we may draw a subsample of size \( n-d \), and there are \( c_n^d \) such possible sub-samples. For each sub-sample of size \( n-d \), we compute the estimator \( T_{n-d}^{(i)} \), where \( i \) stands for the indices of the \( d \) observations not included in this sub-sample. Then, in a manner very similar to (2.5)-(2.6), we may formulate a delete-d jackknifed version of \( T_n \) (denoted by \( T_n^*(d) \)). Similarly, a variant form of (2.8) can be considered as an estimator of \( \sigma^2 \), the asymptotic variance of \( n^{-1/2} (T_n - \theta) \). Actually, we may write

\[
T_n^*(d) = T_n + (n-d)E\{ (T_n - T_{n-d}) \mid C_n \};
\]

\[
V_n^*(d) = (n-d)(n-1) \tau^2 \text{Var}\{ (T_n - T_{n-d}) \mid C_n \},
\]

where \( C_n \) is defined as in before (2.10). This explains the rationality of SRSWOR for delete-d jackknifing, for arbitrary \( d \geq 1 \) [see Sen (1988b)]. There is a variant form of delete-d jackknifing: One may have a resampling plan of \([n/d]\) deletion of distinct sets of \( d \) observations from the basic sample of size \( n \). This scheme has some arbitrariness in the partitioning of the basic sample into \([n/d]\) (overlapping) sub-samples, and hence, we would naturally favor the first scheme where all possible \( c_n^d \) partitionings are considered. For \( d=1 \), these two schemes are the same.

In many situations [viz., the ratio estimator in sampling theory, the sample correlation coefficient, the coefficient of variation etc.], it may be possible to express \( T_n \) as \( g(T_{n1}, \ldots, T_{np}) \), \( p \geq 1 \), where the individual \( T_{nj} \) are averages over independent terms and \( g(\cdot) \) is a smooth (and differentiable) function. In such a case, a (multivariate) Taylor series expansion of \( g(T_{n1}, \ldots, T_{np}) \) yields
\[ T_n = g(\xi_1, \ldots, \xi_p) + \sum_{j=1}^{p} (T_{n,j} - \xi_j) (\partial^2 g(y)) g(y_1, \ldots, y_p) \bigg| y = \xi + 0(||T_n - \xi||^2) , (2.19) \]

where the \( \xi_j \) are the centering constants for the \( T_{n,j} \). Thus, whenever \( n||T_n - \xi||^2 = O(1) \) in \( L_2 \)-norm, the bias of \( T_n \) is also \( O(n^{-1}) \), and the asymptotic mean square of \( n(T_n - \theta)^2 \) is expressible as a quadratic function of the gradients of \( g(.) \) with the covariance matrix of \( n^2(T_n - \xi) \) as the discriminant; this turns out to be a functional of \( F \). The classical delta method consists in estimating this functional by simply estimating the covariance matrix of \( n^2(T_n - \xi) \) from the sample and evaluating \( (\partial^2 g(y)) \) at \( y = T_n \). A similar estimate of the bias (upto the order \( n^{-1} \)) can also be obtained by explicitly writing the last term in (2.19) as a quadratic function involving the second derivatives of \( g(.) \) and using the sample estimates. This delta method is essentially the same as the infinitesimal jackknifing method of Jaeckel (1972), and hence, we shall not discuss the latter.

Looking at (2.19) and earlier decompositions, we may argue that for a given \( n \), the mean square error of \( (T_n - \Theta) \) may be expressed as \( \psi(n) \sigma^2(F) \), where \( \sigma^2(F) \) is a functional of the d.f. \( F \) and \( \psi(n) \) is usually a known function of \( n \). If the form of \( \sigma^2(F) \) is known (while \( F \) is not) then one simple way of estimating it is to take \( \sigma^2(F_n) \) for \( \sigma^2(F) \) where \( F_n \) is the sample d.f. However, in most of the cases (particularly in the nonparametric contexts), the functional form of \( \sigma^2(F) \) [or the actual distribution of \( T_n \)] is not that precisely known. In such a case, bootstrap methods introduced by Efron (1979) may have some advantages over the jackknife methods. Bootstrapping is certainly a clever idea adapted to the advent of modern computing facilities, and it is a resampling scheme in a more meaningful way. We may describe a bootstrap procedure as follows. Recall that \( T_n = T(X_1, \ldots, X_n) \) and \( F_n \) is the sample d.f. corresponding to the population d.f. \( F \). From the sample d.f. \( F_n \), draw with replacement a sample of \( n \) observations (denoted by \( X_1^*, \ldots, X_n^* \)); this is termed a 'bootstrap sample'. Let then \( T_n^* = T(X_1^*, \ldots, X_n^*) \) be the same statistic based on \( X_1^*, \ldots, X_n^* \). (2.20)

Draw (independently) a large number (say, \( M \)) of such bootstrap samples from \( F_n \) and compute the estimators as in (2.20); we denote them by \( T_{n,j}^* \), \( j = 1, \ldots, M \). Further, let \( T_n^* = M^{-1} \sum_{j=1}^{M} T_{n,j}^* \), \( V_n^* = (M-1)^{-1} \sum_{j=1}^{M} (T_{n,j}^* - T_n)^2 \). (2.21)
Then $v_n^*$ is defined as the bootstrap estimator of $E(T_n - \theta)^2$. More generally, we may estimate the unknown d.f. of $n^{1/2}(T_n - \theta)$ by the empirical one based on the set

$$n^{1/2}(T_{n,j} - T_n) = Z_{nj}, \text{ say, for } j = 1, \ldots, M.$$  

(2.22)

We may refer to some recent works of Beran (1988) and Hall (1986, 1988) for deeper results.

The resampling plan in the bootstrap method is thus the simple random sampling with replacement (SRSWR) replicated a large number of times to bestow the capability of estimating the mean square error as well as the sampling distribution of an estimator. In this context, it may not be necessary to assume that (2.13) holds, so that technically bootstrap may have an advantage over jackknifeing (which is based on (2.13)). However, this point should not be overemphasized. A nice example considered by Athreya (1987) shows that bootstrap method may not workout either in some nonregular cases (e.g., a stable law with an index $\alpha < 2$). Further, there remains some arbitrariness in the choice of $M$ [in (2.21)], and the ideal choice of $M$ being indefinitely large is linked to high amount of computer time to generate the bootstrap samples. Thus, it is quite reasonable to attach a cost function to the study of the relative merits and demerits of bootstrapping and jackknifeing methods. I am biased to the latter.

3. STATISTICAL FUNCTIONALS: PRELIMINARY NOTIONS

A statistic $T_n$ expressible as a functional $T(F_n)$ of the sample (empirical) d.f. $F_n$ is called a statistical functional. The population counterpart $\theta$ may likewise be expressed as $T(F)$, the same functional of the true d.f. $F$. Hoffding (1948) used the term 'estimable parameter' or 'regular functional' for $\theta = T(F)$ when there exists a sample statistic $\phi(X_1, \ldots, X_m)$ such that $E_F\phi(X_1, \ldots, X_m) = T(F)$, for all $F$ belonging to a class $F$. Earlier, von Mises (1947) used the term 'differentiable statistical functions' for statistics of the form $\int \cdots \int \phi(x_1, \ldots, x_m) dF_n(x_1) \cdots dF_n(x_m)$, where $m$ is a given positive integer (called the degree) and $\phi(x_1, \ldots, x_m)$ is called the kernel. Actually, statistical functionals may be of more general form than the von Mises' type functionals. Intuitively speaking, one would expand $T(F_n)$ around $T(F)$ in some meaningful way (requiring some sort of differentiability of $T(\cdot)$ at $F$), and hence, such functionals are also termed differentiable statistical functions. Most of the classical
parametric estimators (including the sample moments and maximum likelihood estimators (MLE) of locations and scale) are statistical functionals; so are the sample quantiles, Hoeffding's (1948) U-statistics and von Mises (1947) V-statistics. The advent of robust and nonparametric estimation in the past twenty-five years has produced an abundance of such statistical functionals among which the so called rank based (R-) estimators of location, linear combinations of functions of order statistics (L-estimators) and robust (M-) estimators deserve special mention.

Typically, a von Mises functional with degree 1 is a linear one (expressible as an average of i.i.d.r.v.'s) for which the standard asymptotic theory holds. Keeping this in mind, for nonlinear functionals, the basic goal is to approximate them by suitable linear ones for which the asymptotic theory can be worked out (provided the remainder terms can be handled adequately). This approach has been illustrated in some of the contemporary texts in large sample theory [viz., Serfling (1980) and Pfanzagl (1985)].

Use of statistical functionals in resampling plans has indeed been a popular topic of research in the recent past. The goal is to set a representation of the form

\[ T_n = T(F_n) = T(F) + \int T_1(F;x)d[F_n(x) - F(x)] + \text{remainder term}, \quad (3.1) \]

where \( T_1(F;x) \), the influence function, may generally depend on the unknown \( F \). In this setup, we need to show that the remainder term is negligible to the desired extent and the unknown variance of \( T_1(F;x) \) can be estimated by suitable resampling plans. Thus, we encounter here a nonparametric setup where \( F \) is allowed to belong to a broad class \( F \), and in the usual fashion of nonparametrics, various robustness and (asymptotic) efficiency considerations dominate the scenario. Therefore, there is a genuine need to examine the various resampling plans (discussed in the earlier section) for their basic adaptability as well as inherent robustness preserving aspects. Keeping this basic objective in mind, we first introduce the basic notions on differentiable statistical functionals; the regularity conditions to follow are somewhat more stringent than the parallel ones needed only for the asymptotic normality result; this will be elaborated later on.

Let \( A \) be a topological vector space, and let \( C \) be a class of compact subsets of \( A \) such that every subset consisting of a single point belongs to \( C \). Consider a functional \( T(G) \) defined for \( G \) belonging to \( A \). Also let \( ||G-F|| \) stand for the usual sup-norm. Thus,
recalling that a functional $T(.)$ is Hadamard (or compact)-continuous at $F$ if
\[ |T(G) - T(F)| \to 0 \text{ with } ||G-F|| \to 0 \text{ on } G \in C, \] (3.2)
we may set that a minimal requirement for the (strong) consistency of $T(F_n)$ is the Hadamard continuity of $T(.)$ at $F$ [note that $||F_n - F||$ has all the nice convergence properties]. Let $L(A, B)$ be the set of continuous linear transformations from a topological vector space $A$ to another $B$, and let $A^0$ be an open subset of $A$. A function $T: A^0 \to B$ is said to be Hadamard (or compact) differentiable at $F \in A$, if there exists a $T_F' \in L(A, B)$, such that for any $K \in C$,
\[ \lim_{t \to 0} \{ t^{-1} [ T(F + tJ) - T(F) - T_F'(tJ) ] \} = 0, \] (3.3)
uniformly for $J \in K$; $T_F'$ is called the compact derivative of $T(.)$ at $F$. Note that by virtue of the definition of $L(A, B)$ and $T$, we obtain that if $T(.)$ is Hadamard differentiable at $F$, then, for any $K \in C$,
\[ T(G) = T(F + (G-F)) = T(F) + \int T_1(F;x)d[G(x) - F(x)] + R_1(F;G-F), \] (3.4)
where
\[ |R_1(F;G-F)| = o(||G-F||), \text{ uniformly in } G \in K, \] (3.5)
and $T_1(F;x)$ is the influence function of $T(F_n)$ at $x$. Recall that for $F_n$, by the well known results on the Kolmogorov-Smirnov statistics,
\[ n^{\frac{1}{2}} ||F_n - F|| = O_p(1), \] (3.6)
while,
\[ \int T_1(F;x)d[F_n(x) - F(x)] = n^{-1} \Sigma_{i=1}^n \{ T_1(F;X_i) - ET_1(F;X_i) \}. \] (3.7)
Therefore, by (3.4) through (3.7), we obtain that
\[ n^{\frac{1}{2}} [ T(F_n) - T(F) ] = n^{-\frac{1}{2}} \Sigma_{i=1}^n \{ T_1(F;X_i) - ET_1(F;X_i) \} + o_p(1) \]
\[ + O_p N(0, \sigma_1^2), \] as $n$ increases, (3.8)
where
\[ \sigma_1^2 = \text{Var}(T_1(F;X_1)) = ET_1^2(F;X_1), \] (3.9)
and it is assume that $0 < \sigma_1 < \infty$. Note that $\sigma_1 = \sigma_1(F)$ is itself an (unknown) functional. Thus the first order differentiability of $T(.)$ at $F$ and the positivity (and finiteness) of $\sigma_1$ may suffice for the asymptotic normality in (3.8), but there remains the need to have a consistent estimator of $\sigma_1(F)$, so that the studentized form in (3.8) can be used to draw statistical conclusions on $T(F)$. Depending on the resampling plan we use and the estimator we prescribe, we may need some additional regularity conditions which are introduced below.
A functional $T(.)$ is said to be second order compact (or Hadamard) differentiable at $F$ if for any $K \in C$,

$$T(G) = T(F) + \int T_1(F;x)d[G(x)-F(x)] + \frac{1}{2}\int T_2(F;x,y)d[G(x)-F(x)]d[G(y)-F(y)] + R_2(F; G-F) ,$$

(3.10)

where

$$|R_2(F;G-F)| = o( ||G - F||^2 ) ,$$

(3.11)

the functions $T_1(F;x)$ and $T_2(F;x,y)$ are called the first and second order compact derivatives of $T(.)$ at $F$, and we can always normalize them in such a way that

$$\int T_1(F;x)dF(x) = 0 , \quad T_2(F;x,y) = T_2(F;y,x) ;$$

(3.12)

$$\int T_2(F;x,y)dF(x) = 0 = \int T_2(F;x,y)dF(y) \quad \text{a.e.}$$

(3.13)

The expansion in (3.10) for $G = F_n$ gives a second order representation for the statistical functional $T(F_n)$. It is interesting to note that if $T_1(F;x) = 0$ a.e. then

$$T(F_n) - T(F) = \frac{1}{2}\int T_2(F;x,y)d[F_n(x)-F(x)]d[F_n(y)-F(y)] + o_p(n^{-1}) ,$$

(3.14)

where the first term on the right hand side of (3.14) is a von Mises statistic of degree 2 and it is stationary of order 1 (in the sense of Hoeffding (1948)). Then, from the basic results in Gregory (1977), we obtain that there exists a set of finite or infinite collection of eigenvalues $\{\lambda_k\}$ of $T_2(.)$ corresponding to orthonormal functions $\{\tau_k(.)\}$, such that

$$\int T_2(F;x,y)\tau_k(x)dF(x) = \lambda_k \tau_k(y) \quad \text{a.e.}(F), \quad \forall k \geq 0;$$

(3.14)

$$\int \tau_k(x)\tau_q(x)dF(x) = \delta_{kq} = 1 \quad \text{or 0 acc. as} \quad k = q \quad \text{or not} , \quad k, q \geq 0 .$$

(3.15)

As a result, we have for $T_1(F;x) = 0$ a.e.(F), under (3.10),

$$2n[ T(F_n) - T(F) ] \overset{p}{\to} \sum_{k \geq 0} \lambda_k \left( Z_k^2 - 1 \right) ,$$

(3.16)

where the $Z_k$ are i.i.d.r.v.'s with the standard normal d.f. This exhibits the basic role of influence functions in the asymptotic distribution theory of statistical functionals. This has an important bearing in the effectiveness of a resampling plan. We rewrite $T_1(F;x)$ as

$$T_1(F;x) = \lim_{\varepsilon \to 0} \left\{ e^{-1} \left[ T((1-\varepsilon)F + \varepsilon \delta_x) - T(F) \right] \right\} , \quad x \in E ,$$

(3.17)

where

$$\delta_x(y) = \mathbf{1}(y \leq x) , \quad \text{for} \quad y \in E , \quad x \in E .$$

(3.18)

Thus, we may interpret $T_1(F;x)$ as a measure of sensitivity of $T(.)$ to the single point $x$, and based on this interpretation, functionals with bounded influence functions are
more appealing from the robustness point of view. Recall that the influence function corresponding to the sample mean is $x - \mu$, and hence, it may not be bounded when the support of the d.f. $F$ is the entire line $E$. For the sample median the influence function is $\{2F^{-1}(1/2)\}^{-1} \text{sign}[F(x) - 1/2]$ which is bounded but not smooth at the population median itself. For R-estimator of location based on the classical normal scores statistic, the influence function is $\phi^{-1}(F(x))\int \{f(-f'(y))\phi^{-1}(F(y))dy\}^{-1}$ where $\phi(x)$ stands for the standard normal d.f., so that as $x \to \pm \infty$, $T_1(F;x)$ also $\to \pm \infty$. For M-estimators and L-estimators, generally the score functions are taken as bounded, and in such a case, the influence function is bounded. However, not every functional which is qualitatively robust has a bounded influence function. If $L_F(T_n)$ stands for the probability law of $T_n$ when the true d.f. of $X_1$ is $F$, then the sequence $\{T_n\}$ is said to be qualitatively robust at $F = F_0$ if the sequence $\{L_F(T_n)\}$ is equicontinuous at $F_0$ with respect to the Prokhorov metric $\Pi$ on the space of probability measures; for some of these technicalities, we may refer to Huber (1981, pp. 25-28).]

As a measure of the sensitiveness of the influence function to local shifts, one may consider

$$\gamma(T) = \sup\{ \frac{|T_1(F;x) - T_1(F;y)|}{|x-y|} : x \neq y \}.$$  \hspace{1cm} (3.19)

Again from (local) robustness considerations, one would prefer to have statistical functionals $\{T(F_n)\}$ for which $\gamma(T)$ is finite. For the sample mean, it is so, but for the median (or any quantile), it is not. The interplay of such robustness considerations in resampling plans constitutes an important area of research.

Two important variations deserve mention. First, the functional $T(.)$ may itself depend on $n$ [i.e. $T_n(.)$]. This is typically the case with L-estimators and also in some other robust estimators. The influence function may then depend on $n$ too, and hence, some sort of uniformity conditions need to be imposed along with the ones discussed above. Secondly, one may have a triangular scheme $\{X_{ni}, i \geq 1 ; n \geq 1\}$ of row-wise i.i.d.r.v.'s where $X_{ni}$ has a d.f. $F(n)$ depending on $n$. If $F_{nn}$ stands for the empirical d.f. for the nth row, we may consider a statistical functional of the form $T_n(F_{nn})$, and parallel to (3.4)-(3.5) or (3.12)-(3.13), we may consider a first or second order representation where the compact derivatives are possibly dependent on $n$, and we need to assume the compact differentiability, uniformly in $n \geq n_0$. In such a case, the influence function $T_{n1}(F(n);x)$
may have asymptotically a degenerate distribution, so that the asymptotic normality of $n \frac{1}{2} [ T_n(F_{nn}) - T_n(F(n))]$ may not hold. Thus, in such a case, a more critical appraisal may be needed to probe the effectiveness of the usual resampling plans.

In the rest of this section, we make some general comments on the differentiability aspects of statistical functionals pertinent to further statistical analysis. Earlier works [viz., Kallianpur (1963)] in this context include the formulation based on the so-called Gâteaux derivative [of $T(.)$ at $F$] defined by

$$\lim_{t \to 0} \left\{ t^{-1} [ T((1-t)F + tG) - T(F) ] \right\} = \int T_1(F; x) dG(x), \forall G \in \mathcal{G}; \quad (3.20)$$

where $T_1(F; x)$ may again be identifiable with the influence function, defined in (3.17). However, for the existence of the influence function, the Gâteaux differentiability is not necessary. In robust estimation, although the influence function occupies a central place, the Gâteaux or other related differentiability aspects play a very vital role. In this context, the Fréchet derivative of $T(.)$ [a term commonly arising in differentiation in Banach space] has also been used by a number of workers [viz., Clarke (1983, 1986) and Parr (1985), among others]. The intricacies of such differentiability conditions were examined thoroughly by Reeds (1976), Fernholz (1983) and Clarke (1986), among others. It follows from their studies that the existence of Fréchet derivative implies the same for Hadamard derivative, and in our deeper statistical analysis, these compact derivatives suffice. Hence, as in Sen (1988a), we shall only consider the case of Hadamard derivatives. In this respect, we make a final remark concerning the Hadamard differentiability of $T(.)$ arising in M- and R-estimation of location. Because of their implicit functional forms, in this context, generally one requires that the score function underlying the M- or R-procedure is bounded and smooth [see for example, Fernholz (1983)]. For M-estimators, this is not very stringent, as robustness considerations generally lead to the use of bounded score functions [viz., Huber (1981)]. On the other hand, for R-estimators, such a boundedness condition excludes the important cases of normal scores, log-rank scores and some other unbounded score functions. If our goal is to study only the asymptotic normality result, Hájek (1968) polynomial approximation can be used to include these unbounded scores in the domain (through a sequence of bounded scores). However, this elegant technique may not workout neatly for deeper analysis. We pose this in a more general
mold as follows. Suppose that there exists a sequence \( \{T_r(.); r \geq 1\} \) of statistical functionals, such that (i) for each \( r \), \( T_r(.) \) is first order Hadamard differentiable with an influence function \( T_{r1}(F;x) \), and (ii) \( \sigma_r^2 = E_F T_{r1}^2(F;X_i) < \infty \). Suppose further that there is a functional \( T(.) \), not necessarily Hadamard differentiable, such that (i) as \( r \to \infty \), \( nE_F [T_r(F) - T(F)]^2 \to 0 \), and (ii) \( \lim_{r \to \infty} \sigma_r^2 = \sigma^2 \) exists and is finite. Then it is easy to show that \( n^{1/2}[T(F_n) - T(F)] \) is asymptotically normal with 0 mean and variance \( \sigma^2 \). Next, for each \( r \), for \( T_r(F_n) \), define the pseudovariables as in (2.5) and the jackknifed estimator \( T_n^* \) by (2.6). Show that the conditions stated above are not sufficient to claim that \( n^{1/2}[T(F_n) - T(F)] \) is asymptotically normal (or equivalent to \( n^{1/2}[T(F_n) - T(F)] \), in probability). The main difficulty stems out of the nature of the pseudovariables in (2.5) which require more delicate manipulations, and the Hajek (1968) type projection results require more stringent regularity conditions. Perhaps, this problem illustrates the need to look at the statistical functionals in the context of resampling plans in a deeper way than in the usual study of their asymptotic normality results.

4. STATISTICAL FUNCTIONALS IN RESAMPLING PLANS: AN APPRAISAL

Let us consider now the adaptability of various resampling plans for statistical functionals of general forms. In this context, we would place due emphasis on the robustness aspects of such functionals, and comment on the resampling plans with respect to their extent of preserving robustness in addition to their main objectives.

The first and foremost objective of jackknifing [viz., Quenouille (1956)] was to reduce the bias of an estimator, and this has been explained in (2.7). Let us have a deeper look into this bias reduction business for jackknifing of statistical functionals. Suppose now that a functional \( T(.) \) is second order Hadamard-differentiable at \( F \) [see (3.10)-(3.11)], and define \( T_2^*(G) = \int T_2(G;x,x)dG(x), G \in A \). Then, we may strengthen (2.16) to the following: If \( T_2^*(.) \) is Hadamard-continuous at \( F \), then

\[
(n-1)[T(F_n) - T_n^*] + \frac{1}{n} T_2^*(F) \text{ a.s., as } n \to \infty;
\]

see Parr (1985) and Sen (1988a) for details. This result has two important bearings:

First, it shows the close relation between the original estimator \( T(F_n) \) and its jackknifed version \( T_n^* \), and explains how jackknifing really amounts to the elimination
of the first order bias term without structurally affecting the original estimator. Secondly, it also raises the alarm: If \( T(F_n) \) is itself not a very robust estimator, its jackknifed version \( T_n^* \) will also not be a very robust estimator. Thus, from the robustness point of view, underlying jackknifing of statistical functionals, the choice of a robust initial estimator is very crucial. Actually, it is more. By their very construction [see (2.2) and (2.5)], the pseudovariables are more vulnerable to outliers and error contaminations (than the original \( T(F_n) \)), and hence, the classical jackknifing may not be the right course to preserve or enhance robustness in the resampling scheme. Before we comment on some variant forms of the classical jackknifing, more suitable from the robustness point of view, we may note that (4.1) is restricted to the class of second order Hadamard-differentiable functionals, and such a smooth result may not hold when \( T(.) \) may not belong to this class. A very notable example of such a functional is the sample median (quantile) treated in Example 1 in Section 2. In such a case, the rate of convergence is not \( O(n^{-1}) \) a.s., but \( O(n^{-\gamma}) \) a.s., for some \( \gamma \leq 3/4 \). Thus, the closeness of a statistical functional and its jackknifed version may depend a lot on its underlying differentiability properties. In a somewhat different context, we shall show later on that jackknifing may not be at all suitable for the primary purpose of bias reduction of an estimator. There remains the basic issue of assessing the weak or a.s. convergence of \( n^{1/2}|T(F_n) - T_n^*| \) under minimal conditions on \( T(.) \), and we pose this as one of the open problems. Does the first order Hadamard-differentiability of \( T(.) \) at \( F \) suffice for the convergence of \( n^{1/2}|T(F_n) - T_n^*| \) to 0, in probability/a.s., as \( n \to \infty \)? Note that (3.8) [requiring only the first order Hadamard differentiability of \( T(.) \) at \( F \)] and (4.1) [based on the second order Hadamard differentiability] insure the asymptotic normality of \( n^{1/2}[T_n^* - T(F)] \). Thus, a related question is: Does this asymptotic normality of the classical jackknifed estimator \( T_n^* \) remain in tact sans the second order Hadamard differentiability assumption on \( T(.) \)? Again the example of the sample median (quantile), treated in an elegant way by Ghosh (1971), can be cited as a counterexample to such a claim. Variants of the classical jackknifing, to be considered later on, meet this demand under more general regularity conditions, although they may not have the sharp rate for the bias in (4.1).
Variance estimation by the classical jackknifing method plays an important role in statistical inference [viz., in attaching a confidence interval for \( T(F) \), or testing suitable hypothesis on \( T(F) \)]. In this respect, we may strengthen (2.14) to the following [viz., Parr (1985) and Sen (1988a)]:

If \( T(.) \) is first order Hadamard-differentiable at \( F \) and \( T_1^*(G) = \int T_1^2(G,x) dG(x) \) is Hadamard-continuous at \( F \), then \( V_n^* \), defined by (2.8), converges a.s. to \( \sigma_1^2 \), defined by (3.9), as \( n \to \infty \).

The interesting feature to note in this connection that the first order Hadamard-differentiability of \( T(.) \) practically suffices for the a.s. convergence of the classical jackknifed variance estimator. Even so, there are two major points to ponder in this respect. First, this a.s. convergence result may not hold for the sample median (or quantiles), although the asymptotic normality of \( T(F_n) \) holds. Secondly, looking at (2.2), (2.5) and (2.8), we may gather that from the robustness perspective, all the pseudovariables may lead to even more vulnerability of \( V_n^* \) to outliers and/or error contaminations. Taken together, these two points raise questions regarding the adaptability of the classical jackknifing method in the broad setup of robust estimation of general statistical functionals.

Among the variants of the classical jackknifing, we may consider first the delete-d jackknifing method, mentioned in Section 2. To illustrate the utility of this method, we start with the example of sample quantiles, discussed in Example 1 in Section 2. Let \( X_{n:1}, \ldots, X_{n:n} \) be the sample order statistics for the sample size \( n \), and let \( k_n \) be the nearest integer to \( (n+1)p \) where \( p \in (0,1) \). Then \( T_n = X_{n:k_n} \). We consider a smooth version of the sample quantile for the sample size \( n-1 \), namely,

\[
T_{n-1} = \alpha X_{n-1:k_n-1} + (1-\alpha) X_{n-1:k_n} ; \quad \alpha = n^{-1} k_n .
\]

Then, by definition in (2.5), we have

\[
T_n^{(i)} = \begin{cases} 
\alpha X_{n-1:k_n} + (1-\alpha) X_{n-1:k_n} & \text{for } n-k_n \text{ values of } i, \\
\alpha X_{n:k_n} + (1-\alpha) X_{n:k_n+1} & \text{for } k_n \text{ values of } i, \\
\alpha X_{n:k_n-1} + (1-\alpha) X_{n:k_n+1} & \text{for } 1 \text{ value of } i.
\end{cases}
\]

As such, some standard computations lead us to

\[
T_n^* = T_n + n^{-1}(n-1)[ \alpha(1-\alpha)(X_{n:k_n-1} - X_{n:k_n+1}) + n^{-1} \alpha( X_{n:k_n} - X_{n:k_n-1} ) ] .
\]
At this stage, we may note that whenever the density at the quantile is positive,

\[(n-1)\{X_{n:k_n+1} - X_{n:k_n-1}\}\]

has a nondegenerate limiting distribution, (4.5) so that from (4.4) and (4.5), we conclude that

\[n(T_n - T_n^*) + \alpha(1-\alpha)(n-1)\{X_{n:k_n+1} - X_{n:k_n-1}\} = o_p(1);\]

(4.6)

\[n(T_n - T_n^*)\]

has a nondegenerate limiting distribution, same as that of (4.5) with a scalar factor \(\alpha(1-\alpha)\). (4.7)

This shows that in this case (4.1) does not hold, although we have a similar result involving a stochastic limit for the bias factor \(n(T_n - T_n^*)\). In a similar fashion, it can be shown that in this case, \(V_n^*\), the jackknifed variance estimator, does not converge to \(\sigma_1^2\) \((= p(1-p)/r^2(\xi_p))\). Rather, \(V_n^*\) has a limiting nondegenerate distribution whose centering constant is \(\sigma_1^2\). If instead of choosing \(\alpha = n^{-1}k_n\), we would have chosen a different value (viz., \(\alpha = 0\) or 1), the results would have been quite similar. This explains that the classical jackknifing may not apply to sample quantiles or similar functionals which may not be (second order) Hadamard-differentiable. For such functionals the delete-d jackknifing method may work out better. For every \(i = (i_1,\ldots,i_d)\), based on the set \(S_{n_i,d} = n \setminus i, n = \{1,\ldots,n\}\), we may define \(T_n^{(i)}\), so that the pseudovariables generated by the delete-d jackknifing are defined as

\[T_{n,i}^{(d)} = d^{-1}\{nT_n - (n-d)T_{n_i}^{(i)}\}, i \in I, \text{ the set of } \binom{n}{d} \text{ possible } i.\] (4.8)

Then, the delete-d jackknifed estimator is

\[T_n^{*} = \left(\binom{n}{d}\right)^{-1}\sum_{i \in I} T_{n,i}^{(d)}\]

(4.9)

and, we also introduce the variance function

\[V_n^{*} = \left(\binom{n}{d}\right)^{-1}\sum_{i \in I} (T_{n,i}^{(d)} - T_n^{*})^2.\]

(4.10)

Looking at (2.17) and (2.18), we may verify that

\[T_n^{*}(d) = T_{n,d}^* \quad \text{and} \quad V_n^{*}(d) = d(n-1)(n-d)^{-1}V_{n,d}^*, \quad \text{for all } d \geq 1.\]

(4.11)

The presence of the factor \(d^{-1}\) in (4.8) has an averaging effect, for large values of \(d\). Thus, if we allow \(d = d_n\) to depend on \(n\) in such a way that as \(n\) increases, \(d_n\) also does so, but \(n^{-1}d_n + 0\), then we may carry out an analysis very similar to that in (4.3) to (4.7) and conclude that here (4.1) holds (although with a different form of the \(T_2^*(F)\)) and also \(V_n^{*}(d)\) converges to \(\sigma_1^2\) a.s., as \(n \to \infty\). Although in this specific case, the results follow by using the usual asymptotic properties of sample order statistics,
there is an open question regarding the handling of non-smooth functionals. We may refer
to Shao and Wu (1986) for some interesting observations on such functionals. We may also
pose the following approach based on the celebrated Bahadur (1966) representation of sample
quantiles, generalized to a broader class of (non-linear) estimators. An estimator $M_n$
is said to admit a Bahadur representation (of the first order) if

$$M_n - \theta = n^{-1} \sum_{i=1}^{n} \phi(X_i, \theta) + R_n,$$

where $R_n$ converges to 0 (as $n \to \infty$) at a suitable rate and in a convenient manner (viz.,
in probability/a.s./rth mean, $r > 0$), and $\phi(x, \theta)$ is a suitable score function. Looking at
(4.8) and using the same notations for the $R_n$, we have

$$T_{n,i} = \theta + \frac{1}{n} \sum_{j \in I} \phi(X_j, \theta) + \frac{1}{n} \sum_{j \in I} n R_n - (n-d) R_{n,d}$$

so that

$$T_{n,d}^* = \theta + n^{-1} \sum_{i=1}^{n} \phi(X_i, \theta) + R_{n,d},$$

and hence, the asymptotic bias term for the delete-d jackknifing method can be studied
with the aid of the asymptotic behavior of $n(R_n - R_{n,d}^*)$. Similarly, if

$$\left( \sum_{i \in I} \{ R_{n,i}^* - R_{n,d}^* \} \right)^2 = o_p((n-d)/(d(n-1))),$$

then it follows from (4.11), (4.13), (4.14) and (4.16) that $V_{n,d}^*$ converges in proba-
bility to $\sigma^2 = \text{var}(\phi(X_1; \theta))$, as $n \to \infty$; the corresponding a.s. result will hold if in
(4.16), the right hand side is $o((n-d)/(d(n-1)))$ a.s. This calls for a more indepth
study of the remainder term in (4.12) with special emphasis on the jackknifed version
$R_{n,d}$, when $d$ is large. In this context, it may be noted that

$$R_{n,i}^* - R_{n,d}^* = -(n-d) \sum_{j \in I} \{ R_{n,d}^* - R_{n,d}^* \} \{ R_{n,i}^* - R_{n,i}^* \} \{ R_{n,i}^* - R_{n,i}^* \},$$

so that in the verification of (4.16), we may as well replace it by

$$\left( \sum_{i \in I} \{ R_{n,i}^* - R_{n,d}^* \} \{ R_{n,i}^* - R_{n,d}^* \} \{ R_{n,i}^* - R_{n,d}^* \} \{ R_{n,i}^* - R_{n,d}^* \} \right)^2 = o_p((d(n-d)(n-1)^{-1}).$$

Using the same notation for $C_n$ as in (2.10), we may rewrite (4.18) as

$$\text{Var}( (R_{n,d} - R_n) \mid C_n ) = o_p( (d(n-d)(n-1))^{-1} ).$$

This particular form is of great convenience in a class of situations where the sequence
$
\{ T_n = T(F_n); n > n_0 \}$ can be approximated by a reversed martingale sequence; note
that $\{ (F_n - F); n \geq 1 \}$ is itself a reverse martingale (process), and hence, for
linear $T(.)$, the reverse martingale structure is ensured. A very simple example of a
non-linear $T(.)$ for which $\{T_n\}$ is a reverse martingale is the Hoeffding's (1948)
$U$-statistic. In general, although $T_n$ may not be Hadamard-differentiable, it may be
sufficiently closely approximated by a reversed martingale. In fact, in (4.12),
n^{-1} \sum_{i=1}^{n} \phi(X_i; \theta), n \geq 1, form a reverse martingale, and hence, $R_n$ may as well be
approximated by a reverse martingale. In a variety of situations, $\{R_n\}$ can be chara-
ceterized as a reverse sub (or super-) martingale, and hence, a verification of
(4.19) can easily be made with the aid of the Hájek-Rényi-Chow inequality. For a
general statistical functional, thus, the adaptability of delete-d jackknifing can be
judged by suitable reverse martingale approximations, and the choice of $d = d_n$ may
depend intricately on this order of approximation.

If, however, we are dealing with Hadamard-differentiable $T(.)$, the question arises:
Is it worthy in switching to a delete-d jackknifing method for $d > 1$ ? Note that from
the computational aspect the smaller is the value of $d$, the simpler is the procedure.
On the top of that there is apparently no extra mileage in adapting a delete-d jack-
knifing method, for $d > 1$, when $T(.)$ is Hadamard-differentiable. This is apparent
from the following results [see Sen (1988b)]:

If $d = d_n$ is $o(n^{-1})$, for some $n > 0$ , then parallel to (4.1), we have

\[
(n-1)\{T_n - T_{n,d}\} \rightarrow 1/2 T_2(F) \text{ a.s., as } n \to \infty, \quad \text{and moreover},
\]
\[
V_n^*(d) = V_n^* + o(1) \text{ a.s.} = \sigma_1^2 + o(1) \text{ a.s., as } n \to \infty. \tag{4.20}
\]

Thus, we would advocate the use of delete-d jackknifing instead of the classical one
when $T(.)$ may not be Hadamard-differentiable but may be approximated suitably by a
reverse martingale for which (4.19) may be verified conveniently. From the robustness
aspects (against outliers and/or error contaminations), however, delete-d jackknifing
for some $d > 1$ may serve better.

We now proceed to have a more critical look into this robustness aspect. As has
been mentioned earlier, the $T_{n,i}$ in (2.5) are more vulnerable to outliers or error
contaminations (than the original $X_i$ ), and hence, even $T_n$ were robust, its jacknifed
version may not be so much robust. For this reason, various workers have looked into
the possibility of recombinining the pseudovariables in a more robust manner to obtain
a more robust jackknifed version. Hinkley and Wang (1980) proposed a trimmed mean of these 
T_{n,i} (instead of the ordinary mean \( T_n^* \)) while Parr (1985) considered a variant form of a 
L-functional of these pseudovariables as a jackknifed version. Let us have a look into 
these pseudovariables with a view to incorporating a general functional as a jackknifed 
version of \( T_n \); this is termed a functional jackknifing [Sen (1988a)]. To motivate this,
we define the \( T_{n,i} \) as in (2.5) and denote the related empirical d.f. by
\[
G_n(x) = n^{-1} \sum_{i=1}^{n} I( T_{n,i} \leq x ), \quad x \in \mathbb{E}; \quad n \geq 1.
\]
(4.21)
Note that the jackknifed estimator \( T_n^* \) in (2.6) is given by \( \int x dG_n(x) \) and \( V_n^* \) in (2.8) is 
given by \( n(n-1)^{-1} \{ \int x^2 dG_n(x) - ( \int x dG_n(x) )^2 \} \). For a general statistical functional 
\( T(\cdot) \), the \( T_{n,i} \) are not independent (although they are interchangeable). Moreover,
because of the coefficients \( n \) and \( n-1 \) attached to \( T_n \) and \( T_{n-1}^{(i)} \) respectively, the \( T_{n,i} \) are 
more vulnerable to error contaminations and outliers. As such linear functionals may 
not be very appropriate from the robustness considerations, and we may like to choose 
a general functional \( T^0(\cdot) \) with due emphasis on its robustness aspects. Then
\[
T_n^0 = T^0(G_n), \quad \text{for a suitable } T^0(\cdot) \text{ defined on } D[0,1],
\]
(4.22)
is termed a functional jackknifed estimator of \( T(F) \). Note that \( G_n(x) = P( T_{n,i} \leq x ) \), 
the actual d.f. of the \( T_{n,i} \) may generally depend on \( F \) as well as \( n \). It is not unreasonable to assume that there exists a d.f. \( G_F \), depending on \( F \), such that \( ||G_n(x) - G_F|| \) 
\( \rightarrow 0 \) as \( n \rightarrow \infty \), while \( ||G_n - G_n(0)|| \) \( \rightarrow 0 \) in probability/a.s./rth mean. Hence, if \( T^0(\cdot) \) 
is a smooth functional, \( T_n^0 \) will converge in some mode to \( T(F) \) as \( n \) increases. Thus, in 
order that \( T_n^0(\cdot) \) is a relevant functional for the estimation of \( T(F) \), we must have
\[
T^0(G_F) = T(F), \quad \text{for all } F \text{ belonging to a class } \mathcal{F}.
\]
(4.23)
Actually, we may set \( T(F) = \int x dG_F(x), \quad F \in \mathcal{F} \), so that \( T^0(\cdot) \) may be taken as some 
conventional functional related to the usual location model, and this provides the 
justification for the use of \( R-, M- \) and \( L- \)estimators based on the pseudovariables \( T_{n,i} \).
Since such robust estimators are usually translation-equivariant, we may set another 
side condition. Let \( G(x;a) = G(x-a) \), for \( x,a \in \mathbb{E} \). Then, we assume that
\[
T^0(G(.;a)) = a + T^0(G(.;0)), \quad \text{for every real } a.
\]
(4.24)
Thus, if we define \( G_F^*(x) = G_F(x+T(F)) \), \( x \in \mathbb{E} \), then we may confine ourselves to the 
following class of functional jackknifed estimators of \( T(F) \):
$T^0(.)$ is translation-equivariant with $T^0(G_F^*) = 0$.  \hfill (4.25)

The functionals considered by Hinkley and Wang (1980), Parr (1985), Sen (1988a) and others all satisfy this requirement.

Let us define $T^*_n,i = T_n,i - T_n^*$, $i=1,\ldots,n$, and denote by $G_n^*$ the empirical d.f. of these $T^*_n,i$ (so that $G_n^*(x) = G_n(x+T_n^*)$, for every $x \in E$). Then the empirical d.f. $G_n^*$ and its population counterpart $G_F^*$ play the basic role in the study of the asymptotic properties of $T^0_n$: The weak or strong mode of convergence of $G_n^*$ to $G_F^*$ would imply the same for $T^0_n$, while we may use the weak convergence of $n^{1/2}[G_n^* - G_F^*]$ to derive the asymptotic distribution of $n^{1/2}[T^0_n - T(F)]$. There are two important points to mention in this context. Firstly, the $T^*_n,i$ (for a given $n$) are not independent; they are exchangeable r.v.'s with a marginal d.f. dependent on $n$. Thus, the usual limit theorem for the empirical distributional process may not be strictly applicable here. Secondly, if $T^0(.)$ is not strictly a linear functional (a.e.), then $T_n - T_n^0$ may not be so small that (4.1) holds. This second point is overshadowed by the robustness aspect of the functional jackknifing, and it is possible to choose a robust functional $T^0(.)$ such that $n^{1/2}|T_n - T_n^0| = o_p(1)$; some detailed discussions are made in Sen (1988a). The first point, however, requires more careful considerations. We are possibly in the domain of central limit theorems for perturbed r.v.'s. But, here the perturbations are $O_p(n^{-1/2})$, and hence, a more delicate treatment is needed. Based on a weak convergence result on the normalized difference of the empirical d.f.s of the $T^*_n,i$ and $T_1(F;X_i)$, an alternative proof of the asymptotic normality of the normalized functional jackknifed estimator (for first order Hadamard-differentiable $T^0(.)$) is due to Sen(1988a).

When $T^0(.)$ is not a linear functional, the jackknife estimator $V_n^*$ in (2.8) may not be consistent for the asymptotic variance of $n^{1/2}[T^0_n - T(F)]$; the latter is useful in drawing statistical inference on $T(F)$ through $T^0_n$. A robust estimator of this asymptotic variance based on a two-step jackknifing has been considered in Sen(1988a). Let $G_n$ be the empirical d.f. of the $T_n,i$. Also, let $T_{n-2}^{(ij)}$ be the statistic $(T_n)$ computed from a sample of size $n-2$ (dropping $X_i$, $X_j$ from the base sample), and let

\[ T_{n,i:j} = (n-1)T_{n-1}^{(i)} - (n-2)T_{n-2}^{(ij)}, \text{ for } j=1,\ldots,n \ (j \neq i); \ i=1,\ldots,n. \hfill (4.26) \]

We denote the empirical d.f. of the $T_{n,i:j}$ ($j \neq i = 1,\ldots,n$) by $G_{n-1}^{(i)}$, for $i =$
At the second stage of jackknifing, we consider the pseudovariables generated by the functional jackknifed estimators, and denote them by

\[ Q_{n,i} = n\overline{t}^0(G_n) - (n-1)\overline{t}^0(G_{n-i}), \text{ for } i = 1, \ldots, n \]  \hspace{1cm} (4.27)

Let then \[ \bar{Q}_n = n^{-1} \sum_{i=1}^{n} Q_{n,i} \] and

\[ V_n^{**} = (n-1)^{-1} \sum_{i=1}^{n} (Q_{n,i} - \bar{Q}_n)^2. \]  \hspace{1cm} (4.28)

The asymptotic variance of \( n^{-1}( \overline{t}_n^0 - T(F)) \) is denoted by \( \sigma_o^2 \), and it is assumed that the regularity conditions relating to the asymptotic normality of \( n^{-1}( \overline{t}_n^0 - T(F)) \) hold. Then, \( V_n^{**} \) converges in probability to \( \sigma_o^2 \), as \( n \to \infty \). This result can also be extended to an a.s. convergence result under more stringent regularity conditions. Thus, from the robustness considerations, functional jackknifing provides a satisfactory tool for the estimation of \( T(F) \) as well as its asymptotic variance. However, in this context, it should be mentioned that the first goal of reducing the first order bias term by jackknifing is not totally achieved through functional jackknifing. Rather, to induce more robustness, some sacrifice in this first order bias reduction is allowed here.

We have seen that for the jackknifing methods to have effective control on bias, robustness and asymptotic mean square error, Hadamard differentiability plays an important role. There are some notable problems where this differentiability property may not hold, and jackknifing methods may not serve a very effective role. We cite the following.

Example 2 [ Stein-rule estimation problem]. Let \( X_1, \ldots, X_n \) be \( n \) i.i.d. r. vectors having a multi-normal distribution with mean vector \( \theta \) and covariance matrix \( \Sigma \). The classical maximum likelihood estimator of \( \theta \) is the sample mean vector \( \bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i \). It has been proved by Stein (1956) that excepting in the univariate or bivariate models, the sample mean \( \bar{X}_n \) is not admissible (under quadratic error loss), and there exist alternative estimators (known as the Stein-rule or shrinkage estimators) which dominate \( \bar{X}_n \). Recall that \( \bar{X}_n \) is a statistical functional (vector) and that the Stein-rule provides minimax and admissible estimators, compromising on their unbiasedness. On the other hand, the jackknifing primarily aims to reduce the bias of an estimator without necessarily compromising on its efficacy, and providing at the same time an estimator of the sampling variance of the estimator. Since minimization of the risk is the basic goal in the Stein-rule estimation problems, one may wonder how far the bias reduction objective...
of jackknifing incorporates the dual objectives of minimaxity and risk estimation? To examine this, we consider the simplest shrinkage estimation problem for the multi-normal mean vector \( \mu \) involving a quadratic error loss (with a positive definite (p.d.) \( Q \)). Let

\[
S_n = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n) (X_i - \bar{X}_n)', \quad T_n = n \bar{X}_n S_n^{-1} \bar{X}_n \quad \text{and} \quad d_n = \text{ch}_{\min}(Q S_n).
\] (4.29)

[\text{ch}_{\min}(A) stands for the minimum characteristic root.] We consider the following Stein-rule estimator of \( \hat{\mu} \):

\[
\hat{\mu}_n = \left( I - c_n \frac{T_n}{n} Q^{-1} S_n^{-1} \right) \bar{X}_n \quad \text{as} \quad n \to \infty. \quad (4.30)
\]

Note that both \( \bar{X}_n \) and \( S_n \) are statistical functionals, so that \( \hat{\mu}_n \) is a function of such functionals. As in Sen(1986), we may consider the following:

(i) Restricted jackknifed estimator: Jackknifing applied only to the part \( \bar{X}_n \);

(ii) Semi-restricted jackknifed estimator: Jackknifing applied to both \( \bar{X}_n \) and \( T_n \) (but not to the other component involving \( S_n^{-1} \));

(iii) Unrestricted jackknifed estimator: Jackknifing applied to the entire set.

The first case involves a linear operation, and hence the jackknifed version will be the same as \( \hat{\mu}_n \). Thus, in this case, the jackknifing meets the objectives. However, the case is quite different with (ii) and (iii), both of which involve non-linear operations.

The details have been worked out in Sen (1986), and it was seen that the semi-restricted or unrestricted jackknifing may not generally meet the primary goal of minimization of the risk or its estimation. A closer look into this phenomenon reveals that the functional in (4.30) is not Hadamard-differentiable when \( \mu \) is equal to the pivot \( \mu_0 \) or lies in a close neighborhood of it. Borrowing the same idea, it can be shown that for a general (vector) of statistical functionals arising in the context of Stein-rule estimation problems [viz., Sen (1988d)] jackknifing may not generally meet all the objectives, and this conclusion may easily be extended to the case where the Hadamard-differentiability may not hold. In passing, we may remark that for the Stein-rule estimation problems, if the actual \( \mu \) is away from the assumed pivot then the Stein-rule estimators may not possess any effective dominance over their classical counterparts, and in such a case, the jackknifed versions are all asymptotically equivalent to the classical ones.

Let us now examine the role of bootstrapping in the various problems treated in this section. We borrow the notations from Section 2, and for the bootstrap sample \((X_1^*, \ldots, X_n^*)\), let \( F_n^* \) be the empirical d.f. Then, we may set
\[ \theta = T(F), \ T_n = T(F_n) \text{ and } T_n^* = T(F_n^*). \] (4.31)

Note that \( F \) is assumed to be continuous, \( n^{1/2} || F_n - F || = O_p(1) \) and bootstrapping involves SRSWR [see Section 2]. Hence the weak convergence of \( n^{1/2} [F_n^* - F_n] \) to a tied-down Gaussian function follows by some standard arguments. As such, if \( T(.) \) is Hadamard-differentiable and \( T_1(.; x) \) is Hadamard-continuous in a neighborhood of \( F \), then parallel to (3.8)-(3.9), we have here

\[ n^{1/2} [T_n^* - T_n] \Rightarrow n^{1/2} [T_n - T(F)] \Rightarrow N(0, \sigma_1^2). \] (4.32)

where \( \sigma_1^2 \) is defined by (3.9) and \( \Rightarrow \) means asymptotic equivalence in law. This explains the role of the influence function \( T_1(F; x) \) in the asymptotic normality of a bootstrap functional. Moreover, if we set \( s_n^2 = E_{F_n} \left[ T_1^2(F_n; x_1^*) \right] \), then under mild regularity conditions, \( s_n^2 \rightarrow \sigma_1^2 \) a.s., as \( n \rightarrow \infty \), so that the studentized version \( n^{1/2} [T_n^* - T_n]/s_n \) has also the standard normal distribution for large \( n \). This provides a very useful tool for obtaining a confidence interval for \( T(F) \) based on the estimator \( T_n \) and the cutoff points determined from this studentized d.f. The result extends to the vector case as well as to some other forms of non-normal distributions [viz., Romano (1988)]. Note that the Hadamard-differentiability condition is only a sufficient one, and the asymptotic equivalence in law in (4.32) may hold even without this condition, although the limiting distribution may not then be normal.

Let us go back to the Stein-rule estimation problem treated in Example 2 and examine the role of bootstrapping in this context. For the bootstrap sample \( x_1^*, \ldots, x_n^* \), the mean vector, covariance matrix and the test statistic are denoted by \( \bar{x}_n^* \), \( S_n^* \) and \( T_n^2 \), respectively [compare with (4.29)], and parallel to (4.30), we have then the bootstrap Stein-rule estimator:

\[ \hat{x}_n^{\ast \ast} = \{ I - c_n d_n^*(T_n^2) \}^{-1} Q_{n-1} S_n^* \bar{x}_n^* \] where \( d_n^* = ch_p(QS_n^*) \). (4.33)

When the true parameter \( \theta \) differs from the pivot \( (0) \), asymptotically (as \( n \rightarrow \infty \)), \( n^{1/2} || \hat{\theta}_n^{\ast \ast} - \bar{x}^*_n || \Rightarrow 0 \), and hence, \( \hat{\theta}_n^{\ast \ast} \) shares the same (distributional) properties with the classical bootstrap estimator \( \bar{x}_n^* \). The situation is different when \( \theta \) lies in a Pitman-neighborhood of the pivot. For this situation, we like to draw a comparative picture on jackknifing and bootstrapping.

Offhand, one may like to know whether the bootstrap sampling distribution of
From (4.38), we conclude (on using the asymptotic distributional results on von Mises functionals for first order stationary parameters [c.f. Gregory(1977)]) that

$$n[ T(F_n) - T(F) ] \xrightarrow{N} \sum_{k \geq 0} \lambda_k \left( W_k^2 - 1 \right),$$

(4.40)

where the $W_k$ are i.i.d. r.v.'s with the standard normal d.f., and the $\lambda_k$ are the eigenvalues of $T_2(F;.)$ corresponding to an orthonormal system of functions. For $T_2(F;.)$, we denote these eigenvalues by $\lambda_{k,n}$, $k \geq 0$; note that these $\lambda_{k,n}$ are stochastic in nature. If (i) $T_1(F_n;x) = 0$ a.e. $(F_n)$, and (ii) the series $\sum_{k \geq 0} \lambda_k^2$, $\sum_{k \geq 0} \lambda_{k,n}$ converge (absolutely) and $\lambda_{k,n} \rightarrow \lambda_k$ for each $k(\geq 0)$, then from (4.38)-(4.40), we conclude that the bootstrap distribution of $n[ T(F_n) - T(F) ]$ converges to the asymptotic distribution in (4.40). However, the first condition that $T_1(F_n;x) = 0$ a.e. $(F_n)$ may not be generally true, and hence, this convergence result may not hold in general. Actually, we have

$$T_1(F;x) = T_1(F_n;x) - T_1(F;x) \equiv 0(F_n - F),$$

and hence, the second term on the right-hand side of (4.39) has a comparable stochastic variation too, and as a result, (4.40) may not hold for the bootstrap versions.

At this stage, we may draw a parallel between jackknifing and bootstrapping of statistical functionals. For Hadamard-differentiable statistical functionals, jackknifing provides a consistent estimator of the asymptotic variance, so that when the asymptotic distribution is dictated by this asymptotic variance, we may draw conclusions about this asymptotic distribution with the aid of the variance estimator. However, if the asymptotic distribution is not totally dictated by this asymptotic variance, jackknifing may not provide an avenue to estimate this asymptotic distribution (as is the case with Hoeffding's (1948) U-statistics or von Mises' functionals when the parameter $T(F)$ is stationary of order 1 (or more) at $F$). On the other hand, in the bootstrapping method, whenever the asymptotic distribution is totally specified by its variance, one may use either the convergence of $s_n$ to $\sigma_1$ or even the actual bootstrap sampling distribution to gather an idea of the actual asymptotic distribution. Even when this asymptotic distribution is not a sole functional of the asymptotic variance, bootstrap sampling distribution may be used to estimate this asymptotic distribution. Thus, in this respect, the bootstrap method may have an advantage over the jackknifing methods. On the other hand, the role of bootstrapping for nonnormal asymptotic distribution is not that clear too. Also, on bias reduction, jackknifing has a higher score. Turning then to
robustness aspect, we may conclude that both the jackknifing and bootstrapping methods behave very much alike. For example, if $T(.)$ is (second order) Hadamard differentiable and the influence function $T_1(F;x)$ is possibly unbounded, a lack of robustness in $T_n$ will also be felt equally in $T_n^*$. Thus, in order that a bootstrap method works out robustly, it may be necessary to choose a functional $T(.)$ which is itself robust. In this respect, in the earlier part of this Section, we have stressed the utility of functional jackknifing, and in the same vein, we may as well advocate for functional bootstrapping. Or, in other words, in using a bootstrap procedure, one should pay attention to the proper choice of an initial estimator which should be robust and efficient. If $T(.)$ is not Hadamard-differentiable, we have seen that the jackknifed variance estimator may not behave that well. In such a case, the bootstrap method may workout better. A classical example is the sample median or quantile. Efron (1979,1982) has discussed this problem at length and also left a small point unsupported by mathematical analysis: Does the bootstrap estimate of the asymptotic variance of the sample median converge a.s.? In general, it may not do so. If, however, the underlying density is positive at the median and the d.f. admits of a finite $\alpha$-th absolute moment for some $\alpha > 0$, then the bootstrap variance estimator converges a.s. (as $n \to \infty$) to its population counterpart. This elegant result is due to Ghosh et al. (1984) with supplementary results due to Babu (1986). Since the SRSWR scheme is inherent in the bootstrap method, it is natural to go beyond the normal approximation for the asymptotic distribution. For this, we may refer to Bickel and Freedman (1981) as well as Singh (1981) for earlier works, and further extensions are due to Babu and Singh (1984) and Bhattacharya (1987), among others. Their treatment can be extended to statistical functionals under appropriate differentiability conditions. Development on such Edgeworth type expansions for jackknifing has been much more rudimentary. Thus, we may conclude this section with a note that although bootstrapping boosts of a greater scope of applicability, the jackknifing shares a very common ground, and further anticipated developments should lead to the conclusion that in practice a modified jackknifing procedure (viz., functional or delete-d ones) may possess better robustness properties than the classical bootstrap method.
5. GENERALIZED FUNCTIONALS : HALF-SAMPLING METHODS

In the context of stratified sampling, half-sampling methods have been used for the estimation of the variance of an estimator. We may pose this in a slightly more general framework as follows. Suppose that we have $k (\geq 2)$ populations with unknown distributions $F_1, \ldots, F_k$, respectively. From the $i$th population, a sample of size $n_i$ is drawn, and the corresponding sample or empirical d.f. is denoted by $F_i, n_i$, for $i = 1, \ldots, k$; all these samples are assumed to be independent. Consider a functional

$$T(F) = T(F_1, \ldots, F_k) \quad \text{where} \quad F = (F_1, \ldots, F_k).$$

(5.1)

A natural estimate of $T(F)$ is the sample functional

$$T_n = T(F_1, n_1, \ldots, F_k, n_k).$$

(5.2)

In this setup, the $F_i, n_i$ are all stochastically independent, so that $T_n$ is a functional of $k$ independent sample functions. The generalized U-statistics and von Mises functionals all belong to this class. In the context of the Stein-rule estimation problem [viz., Example 2 in Section 4] we have also considered a more general functional (although in a single sample context). Suppose that $G_1, \ldots, G_r$ are d.f.'s (typically, the marginal ones in the case of vector valued r.v.'s), and let $G_{n1}, \ldots, G_{nr}$ be the corresponding sample d.f.'s (all based on a common sample of size $n$). Then, for a functional $T(G)$, we may consider its natural estimator $T_n = T(G_{n1}, \ldots, G_{nr})$. In fact, we may even combine (5.1) and $T(G)$ to form a more general functional involving $k$ independent sets of d.f.'s where within each set we may have more than one argument. For simplicity, we only consider the case of (5.1) and (5.2).

The concept of compact differentiability laid down in Section 3 extends directly to this vector-argument case. Thus, under the first order compact differentiability, we have on letting $n = n_1 + \ldots + n_k$,

$$T_n = T(F_1, n_1, \ldots, F_k, n_k) = T(F) + \sum_{i=1}^{k} \{ \int T_i(1)(F; x) dF_i, n_i(x) + o(||F_i, n_i - F||) \}
= T(F) + \sum_{i=1}^{k} \sum_{j=1}^{n_i} T_i(1)(F; x_{ij}) + o_p(n_i^{-1/2}).$$

(5.3)

With this linear approximation, the asymptotic normality etc., all follow on standard lines. Here, the $T_i(1), i=1, \ldots, k$, may be termed the component influence functions. To make the asymptotic results meaningful, we make the following assumptions:
Thus, we have
\[ n^k [T_n - T(F)] \xrightarrow{d} N(0, \sum_{i=1}^{k} \rho_i^{-1} \sigma_i^2). \] (5.6)

Therefore, for drawing statistical inference on \( T(F) \), there is a genuine need to estimate the variances \( \sigma_1^2, \ldots, \sigma_k^2 \). Moreover, in (5.3), we have only used the linear approximation, whereby the bias term is \( o(n^{-2}) \). It is quite conceivable that under a second order expansion of \( T(.) \) we can collect the bias term (typically \( O(n^{-1}) \)), and hence, some variants of the usual resampling schemes discussed in Section 2 may be used to reduce this bias term to \( o(n^{-1}) \). Robustness considerations outlined in the earlier sections remain pertinent (to a greater extent) in the current context too.

First, we consider a variant of the classical jackknifing method to reduce this bias of \( T_n \) and to estimate its asymptotic variance too. If in the \( i \)th sample, we delete the \( j \)th observation \( X_{ij} \), the resulting empirical d.f. is denoted by \( F_{i,j,n_i} \), for \( j = 1, \ldots, n_i \) and \( i = 1, \ldots, k \). Let then
\[ T_{n_{i,j}} = n_i T_n - (n_i - 1) T(F_{1,n_1}, \ldots, F_{i-1,n_{i-1}}, F_{i,n_i-1}, F_{i+1,n_{i+1}}, \ldots, F_{k,n_k}), \] (5.7)
for \( j = 1, \ldots, n_i \), \( i = 1, \ldots, k \). From these pseudovariables, we construct
\[ T_{n_{i,j}} = n_i^{-1} \sum_{j=1}^{n_i} T_{n_{i,j}}, \quad \text{for } i = 1, \ldots, k; \] (5.8)
\[ s_{i,n_i}^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (T_{n_{i,j}} - T_{n_{i,j}})^2, \quad \text{for } i = 1, \ldots, k. \] (5.9)

By an expansion (essentially) similar to (3.4) and (3.5), we obtain that
\[ T_{n_{i,j}} = T_n + T_{i}(F_{i,n_i}; X_{ij}) + o(1), \quad \text{uniformly in } j = 1, \ldots, n_i; \quad i = 1, \ldots, k, \] (5.10)
where \( F_n = (F_{1,n_1}, \ldots, F_{k,n_k}) \). As such, we have \( T_{n,i} = T_n + o(1) \), for every \( i = 1, \ldots, k \).

Therefore, from (5.9), we have
\[ s_{i,n_i}^2 = (1-n_i^{-1})^{-1} \left( \int T_{i}^2(F_{i,n_i}; x) dF_i(n_i, x) \right) + o(1), \quad \text{for } i = 1, \ldots, k. \] (5.11)

Note that in (5.10) and (5.11), we have \( o(1) \) with probability one; the derivation is quite similar to (2.30) of Sen (1988) and hence is omitted. As such, assuming the (extended) Hadamard-continuity of the functionals in (5.11) and using the definition in (5.5), we readily obtain that
\[ s_{i,n_i}^2 \rightarrow o_i^2 \quad \text{a.s., as } n \rightarrow \infty, \quad \text{for every } i = 1, \ldots, k. \] (5.12)

If \( T(.) \) admits a second order expansion [similar to (3.10) and (3.11) but extended
to the vector case, then we may as well use the pooled jacknifed estimator
\[ T^*_n = \sum_{i=1}^{k} \frac{n_i}{n} T_{n,i}^* \]  
(5.13)
and verify (4.1). Since the \( F_i, n_i \) are independent, in the second order Hadamard derivatives of \( T(.) \) at \( F \), the mixed ones involving two arguments (say, \( F_i, F_{i'}, i \neq i' \)) will have 0 expectation, and the bias will thus be a sum across the \( k \) pure second derivatives.

Although for bias reduction, the jackknifing serves well, one should be more careful with the robustness aspects, as here multiple d.f.'s induce greater scope for possible departures from model based assumptions and for outliers too. This leads us to advocate functional jackknifing in this case too. The theory runs parallel to that in Sen(1988a) with direct modifications along the lines of (5.7) through (5.13). For not so smooth (generalized) functionals, we may also advocate delete-d jackknifing methods, with \( d \) not so small. For smooth functionals, as in Section 4, there is no need to switch to such delete-d procedures.

Let us now consider the role of bootstrapping methods in this context. We refer to Efron(1982) for some interesting discussions on some simple models, and by virtue of the results presented in Sections 3 and 4, we may easily present parallel results for some general statistical functionals. Denoting the bootstrap samples and empirical d.f.'s by \((X^*_{i1}, \ldots, X^*_{in_i})\) and \(F^*_{n_i}, n_i\) respectively, for \(i=1, \ldots, k\), we have
\[ T^*_n = T(F^*_n) = T(F^*_{1,n_1}, \ldots, F^*_{k,n_k}). \]
(5.14)

Now, given the basic result in (5.6), our concern is to estimate the asymptotic variance in (5.6) by the bootstrap method or more generally to show that the bootstrap sampling distribution of \(\frac{n^2}{n} [ T(F^*_n) - T(F_{-n}) ] \) is the same normal distribution when \( n \) is large.

We may use the same expansion as in (5.3) [but replacing the role of \( F_{-n} \) and \( F \) by \( F^*_n \) and \( F_n \), respectively] and obtain that
\[ T(F^*_n) - T(F_n) = \sum_{i=1}^{k} n_{-i}^{-1} \sum_{j=1}^{n_i} T_i(1)(F_n;X^*_{ij}) + o_p(n^{-1}), \]
(5.15)
where, we have made use of the fact that under (5.4), \(|| F^*_n - F_{-n} || = o_p(n^{-1/2}) \). This representation involving linear functionals with (conditionally, given \( F_n \)) independent summands enables us to prove the desired asymptotic normality result, first under a conditional argument, and then passing on to the usual unconditional limit by showing that the conditional mean and variance remain a.e. stable. In this context, the
Hadamard continuity of the functionals $\int T_{i(1)}^2(F_n,x)dF_{i,n_i}(x)$, $i=1,\ldots,k$, (as has been assumed before) justifies the passage to this limit. This weak convergence result, in turn, enables us to conclude that based on a large number of bootstrap samples, the empirical variance of $n^k(T(F_n^*) - T(F_n))$ is a consistent estimator of the asymptotic variance in (5.6). Thus, for (extended) Hadamard-differentiable statistical functionals, the bootstrap and classical jackknifing methods workout very similarly. They are both subject to the same robustness deficiency aspects, and we may therefore advocate the use of functional jackknifing or functional bootstrapping methods to induce more robustness. The bootstrap method may have some advantages for certain functionals (such as the functions of quantiles of the d.f.'s $F_1,\ldots,F_k$) which are linearly approximable but not necessarily Hadamard-differentiable, although in such a case, delete-d jackknifing methods may workout equally well.

In our discussions, so far, we have limited ourselves to the case where $k$, the number of samples, is fixed and $n_1,\ldots,n_k$ are not small. In the so called half-sample method [viz., Efron (1982, ch. 8)], we may even encounter the case where $k$ is large and the $n_i$ are all small; an extreme case is given by $n_i = 2$, for every $i(=1,\ldots,k)$, so that $n = 2k$. In such a case, (5.4) may not hold, and as such, in the expansion in (5.3) or (5.15), the remainder term may not be $o_p(n^{-1})$, even when the functional $T(.)$ is Hadamard-differentiable (in the extended sense). Efron (1982) has discussed the shortcomings of the bootstrap technique in this context, and the same difficulties are encountered in the general case of statistical functionals. A major reason for these drawbacks stems from the fact that splitting the base sample into two 'half-samples' can be done in a variety of ways (leaving much scope for arbitrariness) and within each subsample of size 2, the estimators of the variability may not be that precise too (adding extra variation in the estimated variance of the statistics). The pitfalls can be largely avoided by an appeal to an intermediate setup where $k$ is allowed to be large, but small compared to $n$, so that the individual $n_i$ are not that small. In principle, the situation then becomes comparable to the delete-d jackknifing method where $d$ is large, but $n^{-1}d$ is small. In such a case, it may also be possible to carryout the asymptotic analysis by an appeal to the usual Bahadur-type represent-
tations for various statistics using a sub-sampling approach. This would eliminate the arbitrariness of the splitting of the base sample into subsamples, and also, the domain of this analysis will be extended to a wider class of statistical functionals which may not be Hadamard-differentiable. For some details, we may refer to Sen (1988c). There is room for further work in this direction, particularly, in the moderate sample size case. Efron (1982) contains a good deal of practical indications for the handling of moderate asymptotics in such resampling plans. In the setup of Hadamard-differentiable functionals, the order of the remainder term in (5.3) or (5.15) may not be enough for instituting some adjustments to make the moderate sample size approximations more adoptable. In this context, the prospects depend a lot on extensive simulation studies, and the bootstrap technique is certainly a very useful tool in this venture.

6. RESAMPLING PLANS FOR ESTIMATION OF FUNCTIONALS OF CONDITIONAL DISTRIBUTIONS

The recent years have witnessed a phenomenal growth of the literature on the estimation of a regression function in a nonparametric setup. This has largely been an asymptotic domain and there remains a genuine need to examine the adaptability of such methods for moderate sample sizes as well. In the context of cross validation methods, Efron (1982) has examined the effectiveness of jackknifing and bootstrap methods, and his findings remain pertinent to a wider class of nonparametric regression estimation problems (with some modifications, of course). We shall find it convenient to introduce here a general nonparametric regression estimation problem and to bring in the relevance of resampling schemes in this context. Again, in our discussions to follow, we shall pay due attention to the robustness aspects and confine our interest to suitable statistical functionals. Estimation of bias constitutes an important task in this study. Moreover, there is a little penalty one has to pay for the estimation of a regression function in a complete nonparametric setup (mainly due to the estimability of a conditional distribution function), and this is reflected in a slower rate of convergence of the nonparametric estimates. Since, typically, we are dealing with multivariate observations, there is a greater scope for outliers or error contaminations to creep in. Further, in multivariate models, for robust estimators, the breakdown points may not behave that smoothly, and hence, there is a greater need to address the issue of the choice of a
suitable functional having good robustness properties. From such considerations, the choice of the usual conditional mean of the dependent variable given the independent variables [as has been mostly advocated in the literature; viz. Stone (1977)] may not appear to be that appropriate, and we shall advocate the use of other functionals which are more robust. M- or L-estimators with bounded influence functions may have good (qualitative) robustness properties. In particular, the choice of a conditional median (or quantile) function seems to be very appropriate, although such a choice may lead to a violation of the basic assumption of Hadamard-differentiability. Nevertheless, Bahadur (1966)-type of representations pave the way for the adaptation of resampling schemes in a natural way.

Let \((X_i, Y_i), i \geq 1\) be a sequence of i.i.d.r.v.'s with a distribution function \(F(x, y)\), defined on the Euclidean space \(\mathbb{E}^P\), for some positive \(p\). Consider the conditional d.f. \(G(y|x)\) of \(Y_i\) given \(X_i = x\), defined for \(y \in \mathbb{E}\) and \(x \in \mathbb{E}^P\). Typically, in a nonparametric regression model, we are interested in the behavior of \(G(.|x)\) treated as a function of \(x \in \mathbb{E}^P\). Generally, a location functional \(T(G(.|x))\) is taken as a suitable measure of the dependence of the conditional d.f. \(G(.|x)\) on \(x\), so that we may formally set

\[
\theta(x) = T(G(.|x)) \text{ as a regression function of } Y \text{ on } X. \tag{6.1}
\]

Recall that \(E[Y|x] = \int ydG(y|x)\) is a particular form of such a functional. We may as well use the \(p\)-th quantile (for some \(p \in (0,1)\)) of \(G(.|x)\) as a suitable regression function; the particular choice of \(p = 1/2\) is appealing and it leads to the conditional median (regression) formulation. If \(F\) were a multivariate normal distribution then \(G(y|x)\) would have been a univariate normal d.f. with mean linear in the components of \(x\) and a constant variance; thus, \(\theta(x)\) would be then a linear function of \(x\), leading to the usual linear regression model. However, when \(G(y|x)\) is not normal, such a linear regression model may not be in tact. In such a case, instead of fitting a linear regression model (on a dubious ground), it may be more meaningful to consider \(\theta(x)\) as a proper regression function. This is the basic motivation for the nonparametric regression models. For an arbitrary \(G\), \(\theta(x)\) in (6.1) may be of quite flexible nature and it seems quite reasonable to assume that \(\theta(x)\) is a smooth function of \(x \in \mathbb{E}^P\). If
G(.|x) is continuous in x (a.e.) and if T(.) is Hadamard-continuous, then it is not difficult to show that \( \theta(x) \) is also continuous in x (a.e.). The more smoothness property we may want to bestow on \( \theta(x) \), the more we may have to assume on the conditional d.f. G(.|x). This leads to a hierarchy of models for nonparametric regressions. Logistic or other forms of nonlinear regression models may also be included in this hierarchy. In that sense, the cross validation models in Efron (1982) and others may also be included in this general setup. It may be a better strategy to impose a minimal amount of smoothness conditions on \( \theta(x) \) and to treat it as a pure nonparametric regression function. In such a formulation, in order to match with the flexibility of \( \theta(x) \), it may be quite appropriate to choose a functional T(.) which is itself not very structured and is robust to minor departures from model based assumptions. In this respect, the conditional median (or quantile) regression function appears to be more suitable than the conditional mean regression function.

For simplicity of presentation, we consider the problem of estimating \( \theta(x) \) at a given point \( x_0 (\in \mathbb{R}^p) \). The solution extends directly to the case where \( x_0 \) is allowed to vary in a compact set \( K \), while the passage to the entire space \( \mathbb{R}^p \) may generally require additional regularity conditions. Given the choice of T(.), the basic problem is to obtain a suitable estimate of the conditional d.f. G(.|x) at \( x = x_0 \). For the latter, usually the \emph{nearest neighbor} and kernel methods can be employed. Either of these methods calls for a scanning of the \( X_i \) in a neighborhood of \( x_0 \), and thereby the effective number of observations on which the estimate of G(.|x) is based may be substantially smaller than n. On the other hand, increasing the width of such a neighborhood may introduce serious bias terms (of quite comparable magnitudes), and hence ceases to serve a good purpose. The choice of such a neighborhood (or the kernel in the kernel method) is therefore of quite delicate nature, and this constitutes the main task in the effective use of such nonparametric estimation procedures. Resampling schemes play a vital role in this context.

Consider a suitable metric d(a,b), \( a,b \in \mathbb{R}^p \), and corresponding to the given point \( x_0 \), define \( Z_i = d(X_i, x_0) \), for \( i = 1,\ldots,n \). Usually the Euclidean distance may be taken for d(.). Note that the \( Z_i \) are all nonnegative random variables, and if \( x_0 \) lies well inside the convex hull generated by \( X_1,\ldots,X_n \), then there is a concentration of the
\( Z_i \) at the left hand end-point \((0)\). Let
\[ 0 \leq Z_1^* \leq \ldots \leq Z_n^* \]
be the order statistics of the \( Z_i \).

Suppose that \( Z_i^* = Z_{S_i} \), for \( i = 1, \ldots, n \), so that \( Y_{S_1}, \ldots, Y_{S_n} \) stand for the induced order statistics; note that although the \( Z_i^* \) are ordered, the \( Y_{S_i} \) need not be so. Next, we consider a sequence \( \{k_n\} \) of positive integers, such that
\[ k_n \to \infty \quad \text{but} \quad n^{-1}k_n \to 0 \quad \text{as} \quad n \to \infty. \]

Then, a neighborhood of \( x_0 \) is demarcated by the subset \( \{Z_i^*: k \leq k_n\} \), and we consider an empirical d.f.
\[ G_n^*(y) = k_n^{-1} \sum_{j \leq k_n} I(Y_{S_j} \leq y), \quad y \in E. \]

It is also possible to rewrite (6.4) as \( \{\sum_{j \leq n} I(Y_j \leq y, Z_j \leq Z_k^*)/\sum_{i \leq n} I(Z_i \leq Z_k^*)\} \), which provides a clearer interpretation of the conditional d.f. based on a neighborhood of \( x_0 \) determined by the order statistics of the \( Z_i \). This also explains the motivation for choosing \( k_n \), such that \( n^{-1}k_n \to 0 \) (so that the diameter of this neighborhood shrinks to \( 0 \) as \( n \to \infty \)). Then, a natural estimator of \( T(G(.|x_0)) = \theta(x_0) \) is given by
\[ \hat{\theta}_n(x_0) = T(G_n^*), \]
where \( G_n^* \) is suitably tailored for \( x_0 \) as well. (6.5)

There are some subtle differences between the cases treated in earlier sections and to be considered now. First, in the classical jackknifing or bootstrap methods, we have essentially a simple random sampling scheme (without or with replacement), whereas the \( Y_{S_j} \) included in the definition of \( G_n^* \) in (6.4) are chosen on the basis of a 'nearest neighborhood' principle with respect to the \( Z_j \) (relative to the point \( x_0 \)). Although, given the \( Z_j^* \), the \( Y_{S_j} \) are conditionally independent, they are not (marginally) identically distributed (in this conditional setup). Also, in the definition of the \( Z_j^* \), we have confined ourselves to the lower end point, so that we are able to include a small neighborhood of \( x_0 \). Choosing a 'too small' neighborhood may entail a smaller value of \( k_n \) which may not yield sufficient accuracy for the estimator \( G_n^* \). On the other hand, choosing a relatively larger width will introduce significant bias in \( G_n^* \) (as an estimator of \( G(.|x_0) \)) which is rather of complicated form. Hence, in the study of the (asymptotic) properties of \( \hat{\theta}_n(x_0) \) in (6.5), we need to pay a very careful attention to the proper choice of \( k_n \). Under (6.3), \( G_n^* \) consistently estimates \( G(.|x_0) \) [granted some smoothness conditions on \( G(.) \)], although at a rate slower than
the weak convergence of \( k_n [ G_n^* - G(x_0) ] \) has been established under quite general regularity conditions in the literature (viz., Bhattacharya and Gangopadhyay (1988) and the references cited there). Thus, it may be quite tempting to incorporate this weak convergence result in the study of the asymptotic behavior of \( k_n \{ \hat{\theta}_n(x_0) - \theta(x_0) \} \). Because of the reasons explained earlier, there is a genuine need to study the asymptotic bias in this context. Even if the functional is very smooth (in the sense of compact differentiability), additional complications may creep in due to the nature of \( G_n^* \).

For these reasons, both jackknifing and bootstrap methods have been proposed by various workers. These resampling plans are, however, not based on the entire base sample, but only on the subset \( \{(Z_j^*, Y_{s_j}^*), j=1,\ldots,k_n\} \). Thus, in a jackknifing method we essentially use a SRSWOR scheme on this subset and generate \( k_n \) subsamples of size \( k_n - 1 \) each. Similarly, in a bootstrap method, we draw (under a SRSWR scheme) a sample of size \( k_n \) from this subset of \( k_n \) base sample observations. Because of the SRSWR structure in the bootstrap method, we essentially deal with a sample (of size \( k_n \)) of i.i.d.r.vectors from a distribution which corresponds to the average distribution of the \( k_n \) observations in the chosen subset. This average distribution, in turn, depends on \( k_n \) (or \( n \)) as well as the basic conditional d.f. of \( Y \) (given \( x \)) in the neighborhood of \( x_0 \). This phenomenon calls for additional smoothness conditions on the conditional d.f. \( G(y|x) \) (in a neighborhood of \( x_0 \)) under which the asymptotic properties of jackknifed and bootstrap estimates can be studied conveniently.

For reasons explained earlier, Bhattacharya and Gangopadhyay (1988) advocated the use of a conditional quantile function for (6.1) and studied its asymptotic properties in an elegant manner. Such a quantile function may not be generally Hadamard-differentiable, and hence, requires a somewhat different approach. Gangopadhyay and Sen (1988) have provided the details of Bahadur-type representations for the bootstrap and the classical empirical d.f.'s, and these enable one to provide a suitable linear approximation to the normalized form \( k_n^\frac{1}{2} [ \hat{\theta}_n(x_0) - \theta(x_0) ] \), which in turn provide the access to the study of the desired asymptotic properties. Bhattacharya and Gangopadhyay (1988) have established the asymptotic normality of a conditional quantile under the condition that \( k_n \to n^\lambda \), for some \( \lambda \leq 4/5 \). However, when \( \lambda = 4/5 \), the bias term arising in the
asymptotic representation (of unknown form) may be quite significant, and there are some technical difficulties in providing a suitable estimate of this bias. For this reason, Gangopadhyay and Sen (1988) have set the limitation that \( \lambda < 4/5 \), and in this setup, the asymptotic bias term is negligible. The situation is quite comparable to the density estimation problem. For the estimation of the marginal density of \( Y \), say, at a given point \( y_0 \), in a kernel method of estimation, it is well known that an optimal rate for the band-width \( b_n \) is \( n^{-1/5} \) (and this corresponds to \( k_n = n^{4/5} \)). However, if we consider the problem of estimation of a conditional density function (in a pure nonparametric framework) then a different optimal order for \( b_n \) (or equivalently, \( k_n \)) exists (under additional smoothness assumptions on the conditional density function). The intricate relationship between the kernel and nearest neighborhood methods for the estimation of a conditional distribution function \( G(y|x_0) \), \( y \in E( and, more generally, a functional of it [as in (6.1)]) \) has been studied in detail in Gangopadhyay and Sen (1988), and it has been shown that such relations can be preserved under jackknifing and bootstrapping too.

Although a conditional quantile function may have some desirable properties (including robustness), there are other smooth functionals which may as well be used in the formulation of a nonparametric regression function (as in (6.1)) and in its estimation. For example, the trimmed (or Winsorized) means or more general L-functionals of a conditional distribution \( G(.|x_0) \) may stand as a good competitor to the conditional median or a quantile. As such, it is conceivable to use functional jackknifing or functional bootstrapping methods for the estimation of a functional as in (6.1). For the same reasons (as have been explained after (6.5)), a study of the rationality and asymptotic properties of such a functional jackknifing or bootstrapping method may require additional regularity conditions. A systematic study of these asymptotics is under progress [Gangopadhyay and Sen (1989)]. The usual limitations of jackknife and bootstrap methods, as have been stressed in Section 4, remain ascribable in such a conditional setup too. However, from robustness considerations, jackknife and bootstrap methods in such a functional estimation of conditional distribution functions appear to be far more attractive than the conventional conditional mean approach, and we shall advocate the use of such robust methods.
7. SOME CONCLUDING REMARKS

We have emphasized on the Hadamard-differentiability of functionals in the context of various resampling plans. There are other important functionals which may not be first or second order Hadamard differentiable, but for which jackknifing or bootstrapping may workout equally well. Hoeffding's (1948) U-statistics of degree \( m(\geq 3) \) provide a good example. For \( m = 2 \), a U-statistic is second order Hadamard differentiable, but for \( m \geq 3 \), it may not be (unless the kernel is bounded). Although in this case, the methodology in Section 4 may not workout properly, it follows from Sen(1977) that a reverse martingale approach works out very well. Thus, the results discussed in this paper can be applied to a wider class of statistics which can be approximated (adequately) by some reverse martingales for which the conditional expectation interpretations in Sen (1977) works. Resampling plans have also been used in more general linear models where the errors may be i.i.d.r.v.'s, but not all the observable r.v.'s are so [viz. Wu(1986) and Shao(1988)]. In such a case, the conditional expectation approach or the Hadamard-differentiability approach may not work out. A heuristic weighted jackknifing method has been proposed by Wu(1986) and Shao(1988) and there remains good scope for further work leading to the rationality and optimality of such weights. Variance component models arising in random or mixed effects models constitute another important area where the resampling plans have a vital role, and in this area there is a genuine need to promote more methodological works. (The current literature is flooded with simulation and numerical studies only). In dealing with non-i.d.r.v.'s, the role of the sample d.f. \( F_n \) is not that clear, so that statistical functionals of \( F_n \) may not appear to be that pertinent. However, in the context of statistical analysis based on 'residuals' the sample d.f. of these residuals convey a lot of information. Thus, what we need is to extend the methodology of statistical functionals for non-i.d. and possibly dependent r.v.'s. This task seems to be well within the reach of the modern statisticians.

A class of functionals, termed generalized L-, M- and R-estimators, introduced by Serfling(1984), deserves mention in the context of resampling plans too. Based on the sample \( X_1, \ldots, X_n \) (of i.i.d.r.v.'s), for some positive integer \( m \), let \( h \) be a function from \( E^m \) to \( E \), and let \( H_F \) be the d.f. of \( h(X_1, \ldots, X_m) \). We are interested in \( T(H_F) \). As
in Sen(1983), we may define an empirical kernel d.f. \( H_n \) by
\[
H_n(x) = \left( \frac{n}{m} \right) \sum_{1 \leq i_1 < \ldots < i_m \leq n} I( h(X_{i_1}, \ldots, X_{i_m}) \leq x ), \quad x \in \mathbb{R}.
\] (7.1)
With this, a natural estimator of \( T(H_F) \) is \( T(H_n) \). For \( m \geq 2 \), the summands in (7.1) are not independent, although they are marginally i.d. Strong convergence of \( ||H_n - H_F|| \) has been established by Sen (1983) by a reverse sub-martingale approach and by Janssen et al. (1984) by an alternative method; these two papers also contain the derivation of the weak convergence of \( n^{1/2} [ H_n - H_F ] \). As such, under the usual smoothness conditions on \( T(.) \), the asymptotic properties of \( T(H_n) \) can be studied very conveniently; we refer to Janssen et al. (1984) for a detailed account of these. Based on the usual robustness and bias-reduction considerations, there is a genuine need for employing a jackknife or bootstrap method in this context too. The asymptotic variance of \( n^{1/2} [ T(H_n) - T(H_F) ] \) is also generally unknown (and depends on the unknown \( F \)), so that a jackknife variance estimator (or its bootstrap counterpart) serves a very useful purpose. It appears that the usual definition of the pseudovariables \( T_{n,i} \) in (2.5) remains in tact, although for the \( T_{n,-1}^{(i)} \), we need to go through the corresponding form of \( H_{n,-1}^{(i)} \) defined as in (7.1). As such, the conditional expectation interpretation in (2.10)-(2.11) and the variance estimator in (2.8), with its conditional expectation interpretation in (2.12) all remain valid. We have no problem in soliciting the general theory discussed in Section 4 for either the classical jackknifing or the bootstrap method in this problem. These resampling plans naturally pave the way for drawing statistical conclusions from the \( T(H_n) \). Again, robustness considerations play a vital role in the formulation of suitable \( T(.) \). As such, functional jackknifing may turn out to be very useful in this context too.

REFERENCES


BABU, G.J. (1988). Bootstrap- a review. (to be published)


