ASYMPTOTICS, NONPARAMETRICS, AND TIME SERIES

edited by

SUBIR GHOSH

University of California, Riverside
Riverside, California

A Tribute to Madan Lal Puri

MARCEL DEKKER, INC. NEW YORK • BASEL

 www.dekker.com

19

Breakdown Theory for Estimators Based on Bootstrap and Other Resampling Schemes

GUTTI JOGESH BABU Pennsylvania State University, University Park, Pennsylvania

1. INTRODUCTION

The computer intensive resampling method called bootstrap has become very popular in the mid-eighties and early nineties. Its success lies in the ease of estimation of the sampling distribution, standard error and confidence intervals, with little or no assumptions about the distribution of the underlying population. Among other desirable properties, it derives its strength from the second-order accuracy in estimating the sampling distributions of a wide class of commonly used statistics. Although the bootstrap is generally applied as a nonparametric tool, there are very few studies on robustness of the method.

In this chapter, the bootstrap method is examined using the notion of breakdown point in robustness, in the context of estimation of the variance of an estimator and of confidence intervals. Since the bootstrap utilizes all the data points, in general, the bootstrap estimator of variance of a statistic is not robust even for robust statistics. An alternative resampling procedure known as the half-sample method is explored, to get around this problem.

Another related area is the study of breakdown points for bootstrap quantiles. The quantiles are essential for deriving the confidence intervals.

Research supported in part by NSA grant MDA904-97-1-0023 and NSF grant DMS-9626189.

Effects of Winsorization before resampling on the breakdown points are discussed.

The bootstrap method has become very popular in statistical inference ever since the appearance of the paper by Efron (1979). Efron's work generated an unprecedented enthusiasm among mathematical statisticians as well as data analysts. In a series of papers, Babu and Singh provided theoretical support to Efron's bootstrap. See Babu (1984, 1986), Babu and Singh (1983, 1984a, 1984b), Singh (1981), and Singh and Babu (1990). See also Babu and Bose (1988), Bose and Babu (1991), and Linder and Babu (1994). The work on the asymptotic theory of the bootstrap in the early eighties resulted in establishing the superiority of the bootstrap approximation for a wide class of statistics. Babu and Bai (1996) developed a unified theory for various resampling procedures including the Bayesian bootstrap. The texts by Efron (1982), Efron and Tibshirani (1993), Hall (1992), Shao and Tu (1995), and the review articles by Babu (1989) and Babu and Rao (1993) are some of the selected references to the literature on bootstrap methodology. In spite of active contributions from numerous researchers on several different aspects of the bootstrap methodology, very little is known about robustness, Huber (1981) of bootstrap estimators.

Standard error or sampling variance and confidence bounds are two widely used measures to assess the accuracy in statistical inference. Jackknife and Bootstrap are two of the most popular methods employed in practice to estimate the variance of an estimator, especially when a closed form expression is not available. Henceforth the term estimation of variance will be used to mean estimation of the variance of the sampling distribution of an estimator. Jackknife and bootstrap methods of estimation of variance are briefly described in the next two sections (Sec. 2 and Sec. 3). Does the bootstrap method lead to robust estimation of variance and/or confidence intervals, at least for robust estimators? This question is examined in detail in this chapter.

One measure of robustness is the celebrated concept of breakdown point (see Sec. 4). Stromberg (1997) discusses the breakdown points of jackknife and bootstrap covariance estimators. It is shown in Sec. 5 that the bootstrap estimator of variance of a robust statistic is not necessarily robust. The present article focuses on non-smooth but robust statistics such as the sample median, where jackknife leads to an inconsistent estimation of variance, while the bootstrap estimator is consistent. (The sample median is considered a non-smooth statistic as it cannot be obtained as a smooth function of the sample mean of multivariate random variables.) An alternative resampling procedure called the half-sample method is explored in Sec. 6, and shown to lead to robust estimation of variance. Babu (1992) investigated the half-sample method and related sub-sample methods in detail. The half-sample method has some advantages over the bootstrap, especially in the case of non-homogeneous populations. In particular, Babu (1992) has established that the half-sample method is robust in estimating the parameters of a linear regression model when the errors are heterogeneous. The half-sample method does not share some of the second-order properties enjoyed by the bootstrap method (see Babu (1992), p. 708). Advantages of the half-sample method in robust estimation of variance are detailed in Sec. 6.

The bootstrap has also become a popular method for estimating confidence intervals, see Babu and Bose (1988), and Hall (1992). Quantiles are essential in deriving confidence intervals. The recent results of Singh (1996) on breakdown points for bootstrap quantiles of robust statistics are discussed in Sec. 7. His suggestions to improve the breakdown points are also briefly discussed.

2. JACKKNIFE ESTIMATION OF VARIANCE

Let $\hat{\theta}_n$ be an estimator of θ based on n iid random vectors $X_1, \ldots, X_n$, i.e., $\hat{\theta}_n = f_n(X_1, \ldots, X_n)$, for some function f_n. Let

$$\hat{\theta}_{n,-i} = f_{n-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$$

be the corresponding recomputed statistic based on all but the i-th observation. The jackknife estimator of the variance of $\hat{\theta}_n$ is then given by

$$\mathrm{Var}_J(\hat{\theta}_n) = \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\theta}_{n,-i} - \hat{\theta}_n)^2.$$

For most statistics, jackknife estimation is consistent, i.e.,

$$\mathrm{Var}_J(\hat{\theta}_n)/\mathrm{Var}(\hat{\theta}_n) \rightarrow 1,$$

as $n \rightarrow \infty$. Hence it is very attractive in practice; more so, due to its computational and conceptual simplicity. However, consistency does not always hold; for example the jackknife method fails for non-smooth statistics, such as the sample median. If $\hat{\theta}_n$ denotes the sample median, then in general,

$$\mathrm{Var}_J(\hat{\theta}_n)/\mathrm{Var}(\hat{\theta}_n) \rightarrow \left(\frac{1}{2}\chi_2^2\right)^2$$

in distribution, where χ_2^2 denotes a chi-square random variable with 2 degrees of freedom (see Efron (1982), Sec. 3.4). So in this case, the jackknife method does not lead to a consistent estimator of the variance.

3. BOOTSTRAP ESTIMATION OF VARIANCE

To describe the bootstrap method, let $X_1, \ldots, X_n$ denote independent random vectors with a common distribution function F. Let $\hat{\theta}_n$ be an estimator of θ based on $X_1, \ldots, X_n$, i.e., $\hat{\theta}_n = f_n(X_1, \ldots, X_n)$, for some function f_n. Suppose $X_1^*, \ldots, X_n^*$ is a sample from the empirical distribution function F_n of $X_1, \ldots, X_n$. That is, $X_1^*, \ldots, X_n^*$ are sampled with replacement from $X_1, \ldots, X_n$. Let $\theta_n^* = f_n(X_1^*, \ldots, X_n^*)$. The bootstrap variance Var*, is simply the conditional variance of θ_n^* under the empirical measure, given the data $X_1, \ldots, X_n$. In general it can be shown that

$$\mathrm{Var}^*(\theta)_n^*/\mathrm{Var}(\hat{\theta}_n) \to 1,$$

as $n \to \infty$. In principle, $\mathrm{Var}^*(\theta_n^*)$ is completely known, as it is a known function of the observations $X_1, \ldots, X_n$.

For many simple statistics, it is possible to evaluate the bootstrap variance analytically, by finding a closed form. For example, if $\hat{\theta}_n$ denotes the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$, then its bootstrap variance is $n^{-2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. In the general case, when there is no such simple closed formula, the bootstrap variance is approximated by the Monte Carlo method as follows.

Using a computer, first generate B bootstrap samples of size n each, by simple random sampling with replacement:

$$\begin{array}{ccc} X_1^{*(1)}, & \ldots, & X_n^{*(1)} \\ X_1^{*(2)}, & \ldots, & X_n^{*(2)} \\ \ldots & \ldots & \ldots \\ X_1^{*(B)}, & \ldots, & X_1^{*(B)}. \end{array}$$

Then compute the corresponding statistics $\theta_i^*; i = 1, \ldots, B$. If B is of the order of $(n \log n)^2$, then in general, the sample variance of $\theta_1^*, \ldots, \theta_B^*$,

$$\mathrm{Var}_B(\theta_n^*) = \frac{1}{B} \sum_{i=1}^{B} (\theta_i^* - \bar{\theta}_B^*)^2,$$

provides a good approximation to the bootstrap variance, where

$$\bar{\theta}_B^* = \frac{1}{B} \sum_{i=1}^{B} \theta_i^*.$$

While the jackknife method is not suitable for estimation of variance of non-smooth statistics such as the sample median or sample quantiles, the bootstrap method provides a conceptually simple way of estimation. This raises the question whether the bootstrap estimator of the variance of a robust statistic is robust. In other words, is $\mathrm{Var}^*(\theta_n^*)$ robust for robust statistics $\hat{\theta}_n$?

4. BREAKDOWN POINT

To answer the above question, the concept of breakdown in robustness, attributable to Hampel (1971, 1974), is introduced here. The breakdown point is perhaps the most widely used measure of robustness in modern statistical literature. In order to describe the concept of breakdown point, let $X_1, \dots, X_n$ have a common distribution, and let $\hat{\theta}_n = f_n(X_1, \dots, X_n)$ be a statistic. Let m be the least number of X_i destabilizing $\hat{\theta}_n$. That is, m is the minimum number of data points that need to be replaced by worst possible outliers to move the statistic beyond any bound. Then the breakdown point is defined as m/n. In the case of the sample mean,

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} X_{(i)},$$

the breakdown point is $1/n$, where $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ is the ordering of the data $X_1, \dots, X_n$. The result holds because replacement of the single data point $X_{(n)}$ by a very large real number destabilizes the estimate $\hat{\theta}_n$ of the mean. Singh (1993) provides a detailed discussion of paradoxes in robustness, in particular of breakdown points. On the other hand, for the sample median $m_n = X_{(r)}$, where

$$\begin{aligned} r &= \frac{n}{2} && \text{if } n \text{ is an even integer} \\ &= \frac{(n+1)}{2} && \text{if } n \text{ is an odd integer,} \end{aligned}$$

the breakdown point is $r/n = \frac{1}{2}$ or $\frac{1}{2} + 1/2n$, depending on whether n is an even or an odd integer.

For example if $n = 5$, increasing the largest two observations without any bound will not affect m_5, but changing three points will. So the breakdown point here is $\frac{3}{5}$. In case $n = 4$, decreasing the smallest observation without any bound will not affect m_4. But if two observations are moved, then m_4 will change, so the breakdown point is $\frac{1}{2}$. Therefore the breakdown point of the sample median is approximately $\frac{1}{2}$, when the sample size n is large.

5. BOOTSTRAP VARIANCE OF MEDIAN

Although the sample median is a robust estimator of location, as indicated by the asymptotically maximum possible value for its breakdown point, its bootstrap variance estimator is not robust. To see this let m_n^* denote the sample median of a sample, $X_1^*, \ldots, X_n^*$, from the empirical distribution F_n of $X_1, \ldots, X_n$. It is easy to see that

$$m_n^* \leq X_{(i)} \quad \text{if and only if} \quad Z_{n,i}^* = \#\{1 \leq j \leq n : X_j^* \leq X_{(i)}\} \geq (n/2),$$

where $X_{(i)}$ is the ith order statistic. If P^* denotes the probability measure induced by the bootstrap sampling scheme, then the distribution of $Z_{n,i}^*$ under P^* is binomial with n and $p = i/n$. Simple integration by parts leads to

$$p_{i,n} = P^*(m_n^* = X_{(i)}) = n\binom{n-1}{r-1}\int_{(i-1)/n}^{i/n} u^{r-1}(1-u)^{n-r}\, du,$$

where r is as defined in Sec. 4, and

$$\begin{aligned}\mathrm{Var}^*(m_n^*) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{(i)} - X_{(j)})^2 p_{i,n}p_{j,n} \\ &\geq (X_{(n)} - X_{(n-1)})^2 p_{n,n}p_{n-1,n} \to \infty,\end{aligned}$$

as $X_{(n)}$ is increased without any bound. Consequently the breakdown point of the bootstrap variance is $1/n$. One reason for this is that, with positive probability, however small, the bootstrap median m_n^* can take the value $X_{(n)}$. The breakdown point does not pay any attention to the probability of occurrence of outliers, however small it may be.

Ghosh et al. (1984) have shown that if the tail of the population distribution is not heavy, i.e., $\mathrm{E}|X_1|^\varepsilon < \infty$, for some $\varepsilon > 0$, then

$$\mathrm{Var}^*(m_n^*)/\mathrm{Var}(m_n) \to 1,$$

for almost all sample sequences. Babu (1986) has improved upon this result by establishing the above result under the weaker assumption of

$$\mathrm{E}(\log(1 + |X_1|)) < \infty.$$

He showed that the condition cannot be relaxed and that the bootstrap variance explodes to infinity as $n \to \infty$, for almost all samples, if

$$\mathrm{E}\left(\frac{\log(1 + |X_1|)}{\log\log(3 + |X_1|)}\right) = \infty.$$

In summary, the bootstrap estimator of variance of the sample median, $\mathrm{Var}^*(m_n^*)$ is consistent but not robust. However, recall that the breakdown point for the sample median is $\frac{1}{2}$. Thus the outliers hold unusually strong influence on the bootstrap variance, even when they have no effect on the statistic itself. Similar conclusions can be drawn for the sample quantiles.

Thus in general, the bootstrap estimator of variance of a robust estimator is not robust.

Now it is natural to ask if there is a resampling procedure that gives robust estimates of variance of robust estimators. The answer is yes, and a method called the half-sample method is discussed in the next Section.

6. HALF-SAMPLE METHOD

Estimation procedures based on random collection of subsamples of a sample have been in use for a long time. For example, Mahalanobis (1946) uses a method under the name of interpenetrating samples. Efron (1979) discusses Hartigan's (1969, 1975) work on subsample methods and compares it with the then newly developed, bootstrap method. He notes the first-order asymptotic equivalence of the two methods in several cases.

The half-sample method consists of sampling half the number of data points without replacement from the original sample, $X_1, \ldots, X_n$ first, and then basing estimation on this sub-collection. For example, if the basic sample drawn is $X_1, \ldots, X_n$, $n = 2r$, then a sample of size r is drawn without replacement from $X_1, \ldots, X_n$. Suppose the resample consists of the data

$$X_{n_1}, \ldots, X_{n_r}, \quad \text{where} \quad 1 \leq n_1 < \cdots < n_r \leq n.$$

Then the half-sample estimator of the median $m_{n,H}$, is the median of the points $X_{n_1}, \ldots, X_{n_r}$. Note that if all the original data points, $X_1, \ldots, X_n$ are distinct, then unlike bootstrap samples, the resampled points $X_{n_1}, \ldots, X_{n_r}$ are also distinct.

To understand the basic difference between the half-sample method and the bootstrap method, let $n = 15$ and let the original data be arranged in an increasing order

$$X_{(1)} < X_{(2)} < \cdots < X_{(14)} < X_{(15)}.$$

The size of the half-sample is $[n/2] = [15/2] = 7$, where $[x]$ denotes the greatest integer not exceeding x. Suppose a particular drawing of the half-sample consists of the seven points

$$X_{(2)}, X_{(3)}, X_{(7)}, X_{(10)}, X_{(11)}, X_{(13)}, X_{(14)}.$$

For this realization, $m_{15,H} = X_{(10)}$. It is trivial to note that, irrespective of the values of the original sample, and for any realization of the half-sample, $m_{15,H}$ can never be equal to $X_{(1)}, X_{(2)}$, or $X_{(3)}$. So the half-sample median avoids the least three values of the original data. This is in sharp contrast to the bootstrap method, where the bootstrap median can take any of the values of the original sample. Consequently, the least three values do not contribute to the estimation of the variance of $m_{15,H}$. But $X_{(4)}$ does contribute to the estimation of the variance, since for some sub-samples of size 7, $m_{15,H}$ can take the value $X_{(4)}$. So moving four values to the extreme would substantially alter the value of $m_{15,H}$. Hence the breakdown point in this case is $\frac{4}{15}$. See Babu (1992) for details on the half-sample method of estimation of the variance of quantiles.

To describe the half-sample distribution P_H of the half-sample median $m_{n,H}$ given the data $X_1, \ldots, X_n$, let

$$h = \frac{1}{2}\left(\left[\frac{n}{2}\right] + 1\right) \quad \text{if} \left[\frac{n}{2}\right] \text{ is odd}$$
$$= \frac{1}{2}\left[\frac{n}{2}\right] \qquad \text{if} \left[\frac{n}{2}\right] \text{ is even.}$$

It is shown in Babu (1992) that the sampling distribution of the half-sample median is given by

$$p_{i,n,H} = P_H(m_{n,H} = X_{(i)} | X_1, \ldots, X_n) = \frac{\binom{i-1}{h-1}\binom{n-i}{[n/2]-h}}{\binom{n}{[n/2]}},$$

if $h \le i \le n + h - [n/2]$ and $p_{i,n,H} = 0$, otherwise. This leads to the half-sample estimator of variance

$$\mathrm{Var}_H(m_{n,H}) = \frac{1}{2} \sum_{i,j=h}^{n+h-[n/2]} (X_{(i)} - X_{(j)})^2 p_{i,n,H} p_{j,n,H}.$$

As in the case of bootstrap estimation, the variance estimator depends only on the sample data points. However it is not as easy to compute as the bootstrap estimator of variance of the sample median, as $p_{i,n,H}$ are difficult to compute when n is large.

Note that the order statistics $X_{(i)}$ for $i < h$ do not influence the variance estimator. On the other hand the hth order statistic $X_{(h)}$ has strong influence on the estimation of variance. Even though for large n, the probability $p_{h,n,H}$ is very small, the variance term $\mathrm{Var}_H(m_{n,H})$ can be made to grow without any bounds by moving the hth order statistic, i.e., by moving h observations in one direction. Hence the breakdown point here is h/n, which approaches $\frac{1}{4}$ as $n \to \infty$. This marked improvement in robustness of the half-sample method over the bootstrap is achieved by sacrificing computational simplicity. In addition, have any other desirable properties been sacrificed? Babu

(1992) shows that the half-sample variance estimator is as efficient as the bootstrap estimator. In fact, it is shown that under very general conditions, the relative efficiency,

$$\frac{\mathrm{E}(\mathrm{Var}^*(m_n^*) - \mathrm{Var}(m_n))}{\mathrm{E}\big(\mathrm{Var}_H(m_{n,H}) - \mathrm{Var}(m_n)\big)} \longrightarrow 1,$$

as $n \to \infty$. This holds even when X_i are not necessarily identically distributed. Recall that m_n^* is the bootstrap median and m_n is the sample median.

One of the main reasons for robustness of the half-sample method is that each observation appears in the resampling scheme at most once. On the other hand, in the case of bootstrap method, a data point can appear more than once in the resample.

The same analysis leads to the low breakdown point $1/n$ of the bootstrap estimator of the variance of a sample quantile (see p. 712 of Babu 1992). On the other hand, the breakdown point of the half-sample estimator of the variance of a p-th sample quantile, $0 < p < 1$, is given by k/n, where

$$\begin{aligned} k &= p\left[\frac{n}{2}\right] && \text{if it is an integer} \\ &= \left[p\left[\frac{n}{2}\right]\right] + 1 && \text{otherwise.} \end{aligned}$$

See Theorem 5.1 of Babu (1992). Consequently the asymptotic breakdown point is $p/2$. Similar situation occurs for robust statistics such as the trimmed mean, and certain M and L estimators.

7. BREAKDOWN POINTS FOR BOOTSTRAP QUANTILES

The bootstrap method is, to a large extent used in the estimation of variance and of confidence intervals. Estimation of bootstrap quantiles of a statistic T_n are essential for computing bootstrap confidence intervals, see Babu and Bose (1988), and Hall (1992). In this Section, the breakdown of bootstrap quantiles, as opposed to the sample quantiles are considered. Singh (1996) establishes the relation between the breakdown point of a statistic T_n and the tth bootstrap quantile. He also suggests improvements for certain L and M estimators via Winsorization. To discuss this further, consider 10 percent trimmed mean T_{20} based on a sample of size $n = 20$, see Singh (1996). Even if the largest of the observations $X_{(20)}$ is quite large, the value of the statistic T_{20} is unaffected, as $X_{(20)}$ does not enter into the computation of T_{20}. If the bootstrap method is used to estimate the confidence interval, the extreme value $X_{(20)}$ could appear in the resampled set one or more times. It is quite conceivable that the bootstrap sample consists of 20 copies of the largest value. Even though the probability of this event is extremely low, the outlier has a very strong influence on the bootstrap trimmed mean T^*_{20}. Due to 5 percent trimming on each side, T^*_{20} is free of $X_{(20)}$ if it appears at most once in the bootstrap sample. Suppose Z denotes the number of times $X_{(20)}$ appears in a bootstrap sample. If $Z > 1$, for a particular bootstrap sample, then $X_{(20)}$ will affect the value of T^*_{20}.

Note that the different bootstrap realizations lead to different values of T^*_{20} and to different value of Z. To analyze the bootstrap quantiles, it is necessary to understand how often T^*_{20} is influenced by the extreme value in the sample. Clearly, under the measure induced by the bootstrap sampling scheme, the random variable Z has a binomial distribution Bin(20, 0.05). Hence $X_{(20)}$ influences $100(1-p)$ percent of all the bootstrap samples, where $p \approx 0.736$ denotes the probability that $Z \leq 1$. For such samples T^*_{20} can be made arbitrarily large by letting $X_{(20)}$ increase without any bound. Consequently, the bootstrap quantile Q^*_t of T^*_{20} will tend to ∞ for all $p < t < 1$. So for such t, a single data point has a vast influence; leading to the upper breakdown point $= 1/20$. That is, at least 5% of the data have to approach ∞ in order to carry the statistic to ∞. If $t \leq p$, then Q^*_t is not affected by any outliers other than those influencing T_{20}. So the upper breakdown point of Q^*_t is at least as large as that of T_{20} for $t \leq p$.

Singh (1996) builds upon these arguments and establishes the following result. Let T_n be a robust statistic, based on a sample of size n, with upper breakdown point b/n, i.e., b is the smallest number of observations that need to grow without any bound in order to force T_n to approach ∞. Let T^*_n denote the bootstrap estimator of T_n, and let

$$Q^*_t = \min\{x : P^*(T^*_n \leq x) \geq t\}$$

denote the tth quantile of the bootstrap distribution. Then the upper breakdown point b_t for Q^*_t is m/n, where

$$m = \min\{j : 1 \leq j \leq n, P(\text{Bin}(n, j/n) \geq b) > 1 - t\}.$$

Singh (1996) also establishes that

$$nb_t - b + \frac{z_t\sqrt{b(n-b)}}{\sqrt{n}}$$

is asymptotically bounded, where z_t is the tth quantile of the standard normal distribution.

Singh (1996) suggests Winsorization prior to bootstrapping to improve the breakdown point of the bootstrap quantiles. To Winsorize a β-fraction of the data from each end, replace all the data points not exceeding the ℓth order statistic $X_{(\ell)}$ by the $(\ell+1)$-th order statistic and all the data larger than

$(n - \ell + 1)$-th order statistic by the $(n - \ell)$-th order statistic, where $\ell = [n\beta]$, i.e.,

$$X_i^0 = \begin{cases} X_{(\ell+1)} & \text{if} \quad X_i \leq X(\ell) \\ X_{(n-\ell)} & \text{if} \quad X_i \geq X(n - \ell + 1) \, . \\ X_i & \text{otherwise} \end{cases}$$

Resampling from the Winsorized data produces bootstrap samples that are free of outliers. He establishes that the upper breakdown point b_t^* for Q_t^* is given by

$$b_t^* = \max\left(b_t, \frac{([n\beta] + 1)}{n}\right)$$

for $(1{-}2\alpha)100$ percent-trimmed means, where $0 < \beta < \alpha < \frac{1}{2}$. For additional details on breakdown theory for bootstrap quantiles and the effects of Winsorization, see Singh (1996).

8. CONCLUDING REMARKS

The half-sample estimator of the variance of a sample quantile is shown to have a very high breakdown point. This property is not shared by the bootstrap estimator. Extensions of these results to other robust statistics are under investigation and will be reported elsewhere.

Singh's (1996) general formula for the computation of the breakdown point, for a bootstrap quantile of a statistic, is briefly discussed. For a class of L and M estimators, Winsorization before bootstrapping seems to improve the breakdown value. A detailed account of the current investigations by Singh and his colleagues will be reported elsewhere.

Acknowledgments

The author would like to thank the referees for their valuable remarks which led to a considerable improvement of the chapter.

9. REFERENCES

Babu, G. J., Bootstrapping statistics with linear combinations of chi-squares as weak limit. *Sankhyā, Series A 46*: 85–93, 1984.

Babu, G. J., A note on bootstrapping the variance of the sample quantile. *Ann. Inst. Statist. Math., A 38*: 439–443, 1986.

Babu, G. J., Applications of Edgeworth expansions to bootstrap—a review. In: Dodge Y, ed. Statistical Data Analysis and Inference. Amsterdam: Elsevier Science Publishers B. V., 1989, pp. 223–237.

Babu, G. J., Subsample and half-sample methods. *Ann. Inst. Statist. Math. 44*: 703–720, 1992.

Babu, G. J. and Z. D. Bai. Mixtures of global and local Edgeworth expansions and their applications. *J. Mult. Analysis 59*: 282–307, 1996.

Babu, G. J. and A. Bose. Bootstrap confidence intervals. *Statistics & Probab. Letters 7*: 151–160, 1988.

Babu, G. J. and C. R. Rao. Bootstrap Methodology. In: Rao C. R., ed. Handbook of Statistics 9, "Computational Statistics." Amsterdam: Elsevier Science Publishers B. V., 1993, pp. 627–659.

Babu, G. J. and K. Singh. Inference on means using the bootstrap. *Ann. Statist. 11*: 999–1003, 1983.

Babu, G. J. and K. Singh. Asymptotic representations related to jackknifing and bootstrapping *L*-statistics. *Sankhyā, Series A 46*: 195–206, 1984a.

Babu, G. J. and K. Singh. On one term Edgeworth correction by Efron's bootstrap. *Sankhyā, Series A 46*: 219–232, 1984b.

Bose, A. and G. J. Babu. Accuracy of the bootstrap approximation. *Probab. Theory and Related Fields 90*: 301–316, 1991.

Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. of Statis. 7*: 1–26, 1979.

Efron, B. The jackknife, the bootstrap, and other resampling plans. CBMS-NSF Regional Conference Series in Applied Mathematics, Volume 38, Philadelphia: SIAM, 1982.

Efron, B. and R. J. Tibshirani. An Introduction to the bootstrap. New York: Chapman & Hall, 1993.

Ghosh, M. W. Parr, K. Singh and G. J. Babu. A note on bootstrapping the sample median. *Ann. of Statist. 12*: 1130–1135, 1984.

Hall, P. The bootstrap and Edgeworth expansion. New York: Springer-Verlag, 1992.

Hampel, F. R. A general qualitative definition of robustness. *Ann. of Math. Stat. 42*: 1887–1896, 1971.

Hampel, F. R. The influence curve and its role in robust estimation. *J Amer. Statist. Assoc. 69*: 383–393, 1974.

Hartigan, J. A. Using subsample values as typical values, *J. Amer. Stat. Assoc. 64*: 1303–1317, 1969.

Hartigan, J. A. Necessary and sufficient conditions for asymptotic joint normality of a statistic and its subsample values. *Ann. of Statist. 3*: 573–580, 1975.

Huber, P. Robust Statistics. New York: John Wiley, 1981.

Linder, E. and G. J. Babu. Bootstrapping the linear functional relationship with known error variance ratio. *Scand. J. Statist. 21*: 21–39, 1994.

Mahalanobis, P. C. Report on the Bihar crop survey: rabi season 1943–1944. *Sankhyā, Series A 7*: 269-280, 1946.

Shao, J. and D. Tu. Jackknife and bootstrap. New York: Springer-Verlag, 1995.

Singh, K. On the asymptotic accuracy of Efron's bootstrap. *Ann. of Statist. 9*: 1187–1195, 1981.

Singh, K. Paradoxes in Robustness. In: Ghosh, J. K., Mitra, S. K., Parthasarathy, K. R. and Prakasarao, B. L., ed. A Raghu Raj Bahadur Festschrift. New Delhi: Wiley Eastern Ltd. 1993, pp. 531–537.

Singh, K. Breakdown Theory for bootstrap quantiles. Technical Report #96-015, Department of Statistics, Rutgers University, New Brunswick, NJ, 1996.

Singh, K. and G. J. Babu. On Asymptotic optimality of the bootstrap. *Scand. J. Statist. 17*: 1–9, 1990.

Stromberg, A. J. Robust covariance estimates based on resampling. *J. Stat. Plan. and Inf. 57*: 321-334, 1997.