Comparison of Bootstrap and Jacknife Statistical Procedures

Amanuel Gobena
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports
Part of the Mathematics Commons, and the Statistics and Probability Commons

Recommended Citation
https://digitalcommons.usu.edu/gradreports/1251

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact dylan.burns@usu.edu.
COMPARISON OF
BOOTSTRAP AND JACKNIFE STATISTICAL PROCEDURES

by

Amanuel Gobena

A report submitted in partial fulfillment
of the requirements for the degree
of
MASTER OF SCIENCE
in
Applied Statistics
(Plan B)

UTAH STATE UNIVERSITY
Logan, Utah
1985
ACKNOWLEDGEMENT

I would like to especially thank my major professor, Dr. Ronald V. Canfield, for the many hours of precious help he extended to me during the preparation of this report and to all faculty members of the Applied Statistics Department, as a professor, advisor and a friend made my stay in Utah State University both enjoyable and educational. Moreover, the moral and financial support that have been given to me remains unforgettable in my future career. In addition, I would sincerely like to appreciate and thank Mrs. Roseann Kunz, secretary of the department, with her patience and effort, and with all her smiling and good character being able to help me in typing out this report.

I dedicate this work to my beloved wife, Tamenech Belama, for encouragement, understanding, and sacrifices, and to my parents who gave me the opportunity to go to school from the herdsman and to my brother who took all the responsibilities of supporting me through the school.

I also thank God for this opportunity to serve him. I pray that I can always share the love that I have received during the past years.

Amanuel Gobena
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>i</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>iii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>BOOTSTRAP</td>
<td>2</td>
</tr>
<tr>
<td>Ordinary Bootstrap</td>
<td>4</td>
</tr>
<tr>
<td>Randomized Bootstrap</td>
<td>6</td>
</tr>
<tr>
<td>Double Bootstrap</td>
<td>10</td>
</tr>
<tr>
<td>JACKNIFE</td>
<td>10</td>
</tr>
<tr>
<td>Procedure</td>
<td>10</td>
</tr>
<tr>
<td>Comparisons</td>
<td>13</td>
</tr>
<tr>
<td>Bootstrap in Relation to Median</td>
<td>17</td>
</tr>
<tr>
<td>MEDIAN FOR THE JACKNIFE</td>
<td>19</td>
</tr>
<tr>
<td>Confidence Interval</td>
<td>20</td>
</tr>
<tr>
<td>APPLICATIONS</td>
<td>24</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>24</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>25</td>
</tr>
<tr>
<td>APPENDIXES</td>
<td>26</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Comparisons of Bootstrap, Jacknife, Delta method, and normal theory to the true value</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>Estimates of Standard deviations for Switzer's adaptive trimmed mean using the jacknife and the bootstrap</td>
<td>16</td>
</tr>
</tbody>
</table>
ABSTRACT
Comparison of Bootstrap and Jacknife Statistical Procedures
by
Amanuel Gobena, Master of Science
Utah State University, 1985

Major Professor: Dr. Ronald V. Canfield
Department: Applied Statistics

This report compares the bootstrapping to jacknifing statistical procedures in terms of bias, confidence interval and estimation of median. Related literature have been reviewed. A bootstrap allows a researcher to get an approximation to the distribution of possibly complicated statistical summaries. It is based on random sampling with replacement from experimental units. Jacknife has also been in operation prior to bootstrapping statistical procedure. The jacknife divides the data into subgroups and obtains partial estimates of these subgroups by omitting one subgroup at a time. When both of these statistical resampling procedures are compared the bootstrap has less bias, more accurate confidence interval and better estimation of the median than the jacknife.

(30 pages)
INTRODUCTION

Efron (1979) introduced the bootstrap resampling technique and since then, a considerable body of literature has been developed around the technique. A monograph published by Efron (1982) provides an excellent account of the method and how it can be used. A bootstrap procedure allows one to get an approximation to the distribution of possibly complicated statistical summaries. Formally, the bootstrap distribution of a statistic is the true underlying distribution (i.e., the true distribution) of the statistic. It is based on random sampling with replacement from experimental units. In some situations, one may construct the bootstrap distribution exactly but, in general, a single approximation is used. This paper introduces the general procedure of bootstrap, explains the kinds of bootstrap in brief and compares it to that of jackknife. Extensions of the method of determining the estimates of variance (bootstrap and jackknife) are compared from the literature on the basis of bias, estimates of variance, median and placement of confidence intervals. Procedural approaches are also discussed for both resampling techniques.

Therefore, the purpose of this report is:

(a) To concisely describe the above mentioned methods.
(b) To show how these methods were derived from the same basic idea and,

(c) To draw specific connections between these techniques.

A. Bootstrap

Estimation of bias, variance and the more general measures of error (e.g. standard errors, mean square errors, etc.) have been in operation since the beginning of the era. Recently, however, useful analytical techniques have been derived.

Let $X_1, X_2, \ldots, X_n \sim F$ be independently and identically distributed (iid) with sample size $n$ from the random variable $X$ with unknown $F$ on the real line. Sample averages provide an estimate of the expectation of $X$. However, the data set also provides more information than that used to estimate the statistics.

The usual parametric formula for estimating the standard deviation of the means and standard deviation estimates is:

$$\hat{\sigma} = \left[ \frac{1}{n(n-1)} \sum (X_i - \bar{X})^2 \right]^{1/2}$$

But a problem arises with other statistics, such as a median, since no formula exists for estimation of the standard deviation of the median. Estimating the variance
of the small sample median have been elucidated by Maritz and Jarrett (1978).

1. Take a random sample with replacement from the observed experimental units.

2. Compute and store the statistic based on the new sample.

3. Iterate Steps 1 and 2 N times. No hard and fast rules are available for deciding how many iterations are required; but, in general, 100 to 500 iterations are used, and these variations depend on the objectives of the researcher. The random samples are usually of the same size as the number of experimental units.

4. The distribution of the "bootstrap" values computed from the resampled data. Then, following the above instructions the distribution of the statistic will be approximated.

5. Note that this distribution can be used for many purposes. For example, an approximate standard error of the statistic $X$ is based on the bootstrap distribution of $X$ is the square root of

$$
\sqrt{\frac{1}{B} \sum_{i=1}^{B} (X_i - \bar{X})^2},
$$

where $B$ is the number of iterations, $X_i$ is the value of the statistic from the $i$th iteration and
\[ \bar{X} \] is the average of the values from the iterations.

6. A percentile method (Efron 1982) can be used to develop approximate confidence intervals. To construct an appropriate 80% confidence interval by first constructing a histogram of the \( X_i \)'s, find the points \( a, \beta \) on the base of the histogram so that 10% of the bootstrapped values are to the left of \( a \) and 10% are to the right of \( \beta \); then the interval \([a, \beta]\) is an approximate 80% confidence interval for the statistic. We have to note, however that because bootstrap confidence intervals are suspected to have slight biases, formal significance testing based on these intervals is not recommended.

Efron (1979) has clearly defined three kinds of bootstrapping. These are:

(a) The ordinary bootstrap.

(b) The simple randomized bootstrap.

(c) The double randomized bootstrap.

The Ordinary Bootstrap

Given \( X_1, X_2, \ldots, X_n \) iid \( F \). Let \( \hat{\phi}(X_1, X_2, \ldots, X_n) \) be the statistic, and \( \hat{\phi}(\hat{\phi}) \) be its standard error. However, \( \hat{\phi}(\hat{\phi}) \) depends on:
(a) on the sample size \( n \),
(b) the form of \( p \).
(c) \( F \) which is unknown

Thus, \( \hat{\sigma}_{\text{Boot}} = \sigma(F) \) where \( \hat{F} \) is the empirical probability distribution putting mass \( 1/n \) on each observed value.

\( \hat{F} \) mass \( 1/n \) on \( X_i \), where \( i = 1, 2, \ldots n \) and the \( \hat{\sigma}_{\text{Boot}} \) is the non-parametric maximum likelihood estimate of the \( \sigma(F) \). The following procedure is used by Efron (1977), (1981), (1982).

(a) The function \( \sigma(\hat{F}) \) must be evaluated by the Monte Carlo algorithm.

(b) Construct the sample probability distribution \( \hat{F} \), putting mass \( 1/n \) at each point \( X_1, X_2, \ldots X_n \).

(c) With \( \hat{F} \) fixed a random sample (with replacement) of size \( n \) from \( \hat{F} \) are drawn (i.e. \( X_i^* = X_i, X_n^* = (\hat{F}) \) where \( i = 1, 2, \ldots n \) ) and this is called bootstrap sample \( X' = (X_1^*, X_2^*, \ldots X_n^*) \).

(d) Step C is repeated \( B \) (where \( B \) is large) \( B \) times to obtain bootstrap replications \( (\hat{\rho}(1), \hat{\rho}(2) \ldots \hat{\rho}(B)) \) then the estimate

\[
\sigma_{\text{Boot}} = \Sigma_{B=1}^{B} \left[ \left( \Sigma_{b=1}^{B} (\hat{\rho}(b) - \mu)^2 \right) \right]^{1/2} \]

where \( \mu = \Sigma_{b=1}^{B} \hat{\rho} \).

We have to note that \( B \) is assumed to be large.

Thus, the bootstrap estimate \( \sigma_{\text{Boot}}(\hat{\rho}) \) is simply the estimated standard deviation of the quantity of interest, \( X_1 \ldots X_n \), if the unknown distribution \( F \) is taken to the
empirical distribution $F$. Theoretical calculation of $\sigma_B(\hat{\theta})$ is impossible, but Monte Carlo simulation yields a quick approximation.

**Randomized Bootstrap**

The randomized bootstrap is a particularly simple variant of the ordinary bootstrap appropriate when $y$ is dichotomous. It is quite often used when constructing a prediction rule on the basis of some data, and when it is desired to estimate the error rate of the prediction rule in classifying future observations. It is usually related to cross validation (which provides a nearly unbiased estimate when the original data is used).

Before explaining the procedure of the randomized bootstrap it remains justifiable to clear some of the terminologies as used by Efron (1983). To have some kind of prediction, the statistician is always faced with a set of cases $X_1, X_2, ..., X_n$ which are in general called the training set $X$. Each of these cases consists of two parts $X_i = (t_i, y_i)$ where

- $t_i$ = a vector of predictors and
- $y_i$ = is a response variable

as an example, $t_i$ can describe corn height, disease resistance, insect resistance, weight of above ground biomass, and so on, where $y_i$ might indicate the yield
of the corn. On the basis of a training set, the prediction rule \( n(t, X) \) is constructed.

The intention is to use \((t_0, X)\) to predict a future unobserved response \( y_0 \) on the basis of its predictor vector \( t_0 \). Therefore, the main concern is in the situation where \( y_i \) is a dichotomy, such as "the corn yielded high" or 'yielded low' and the prediction \( n_i = n(t_i, X) \) is likewise dichotomous. Therefore, let

\[
Q[y_i, n_i] = 0 \text{ if } n_i = y_i \\
= 1 \text{ if } n_i \neq y_i
\]

The idea of the randomized bootstrap is to assign some probability (Mass) to the complimentary points \((t_i, y_i^c)\) where \( y_i^c = 1 - y_i \).

Given then, the training set \( X \), we have in mind some way of assigning probabilities to all \( 2n \) points \((t_i, y_i), (t_i, y_i^c)\). For example assign probability on \((t_i, y_i) = 1/n \ \pi_i(y, X) \) where \( \pi_i(y, X) + \pi_i(y^c, X) = 1 \). This last condition means that \((t_i, y_i)\) and \((t_i, y_i^c)\) are assigned total probability \( 1/n \), as with the ordinary bootstrap. Where does then the randomization come in practice. Let \( \mathbb{P}(\text{Rand}) \) be the distribution on the \( 2n \) points given above. Then the randomized bootstrap estimate of \( w \) (where \( w \) being an expectation) is

\[
w(\text{Rand}) = E^* \ \text{op}(X^*, \mathbb{P}(\text{Rand}))
\]

where the training set

\[
X^* = (X_1^*, X_2^*, \ldots, X_n^*), \text{ each component } \mathbb{F}(\text{Rand}) \text{ and } E^*
\]
indicates expectation with respect to the bootstrap
distribution and 'op' being short for optimism which is some
kind of measure of error rate (see Efron (1983)).

If

\[ N_{i,y}^* = \# \{ X_j^* = (t_i, y) \} \]

and

\[ N_i^* = N_{i0}^* + N_{i1} \]

it is possible to express

\[ \text{op}(X^*, \hat{F}(\text{Rand})) \]

as a sum of \( n \) terms,

\[ \text{op}^* = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{2\pi_i(y_i, X) - 1}{\pi_i^*(y_i, X^*) - \pi_i^*(y_i, X)} \right] \]

\[ X \cdot Q[y_i, n(t_i, X^*)] + [\pi_i(y_i, X^*) - N_i^*] \]

(See Efron 1982, 1983)

This reduces the expression \( X_1, X_2 \ldots X_n \overset{iid}{\sim} F. \)

\( \tilde{\pi}_i(y_i, X) = 1 \) and \( \pi_i(y_i, y) = 0. \) The (-) gives the expression.

\[ w(\text{Rand}) = E^*\left\{ \frac{1}{n} \left[ (2\pi_i(y_i, X) - 1) - (\pi_i^*(y_i, X) - \pi_i^*(y_i, X)) \right] \right\} \]

\( X \cdot Q[y, n(t_i, X^*)] \) (Miller, 1974 by

Therefore, both \( \text{op} \) and \( w(\text{Rand}) \) remain valid if \( y_i \) is
replaced by 1 and \( y_i^C \) is replaced by 0 everywhere in the
equation. Thus, most often in dichotomous prediction
problems, the prediction never provides a probability
assessment \( \pi_i(y_i, X) \) as well as a specific prediction
\( n_1(t_i, X). \) For, example Fisher's estimated linear
discriminant is naturally associated with the probability assessment. 

\[ \pi_i(l, X) = \frac{1}{[1 + \exp - (\zeta_i + t_i \hat{\beta})]} \]

and this is an obvious adhoc component to the choice of the numbers .1,...9 for the two randomized bootstraps. In theory the statistician could make a subjective assessment of the uncertainty in each prediction \( \eta(t_i, \gamma) \), in order to assign \( \pi_i(y_i, X) \) and \( \pi_i(\bar{y}_i, X) \). In the sampling experiments the exact assignments seemed less important than keeping them away from 0 and 1. In particular the simple method \( \pi_i(y_i, X) = .9, \pi_i(y_i, X) = .1 \) causes little bias (and as a matter of fact help to correct the bias in the ordinary bootstrap) and gives almost as much improvement as the more complicated method based on

\[ \pi_i(l, X) = \frac{1}{[1 + \exp - (\zeta_i + t_i \hat{\beta})]} \]

It is obvious that \( \hat{F} \) can be a poor estimate of \( F \), particularly if we know that \( F \) is smooth. Using \( \hat{F}(\text{Rand}) \) in place of \( F \) is a form of smoothing. The smoothing is carried out entirely in the y direction. This is handy, since in real applications t may be very complicated, having high dimensionality, censored components, missing values, qualitative and quantitative components, etc.
Double Bootstrap

This is bootstrapping the bootstrap. It automatically corrects the bias of the ordinary bootstrap without increasing the MSE of estimation.

B. The Jacknife:

The jacknife is an extension of an idea due to Quenouille (1949) and is designed to reduce the bias of an estimator. Suppose we have a sample of N independent observations, each from the same distribution which depends on an unknown parameter \( \theta \). Assume that we have a general method for estimating \( \theta \) and let \( \hat{\theta} \) denote this estimator based on all N observations.

Procedure

1. Divide the data into n groups of size K. Let \( \hat{\theta}_i \), = 1, ... n, denote the estimator of \( \theta \) obtained by deleting the ith group and estimating from the remaining \((n - 1)k\) observations.

2. Define \( \tilde{\theta}_i = n\hat{\theta}-(n-1)\hat{\theta}_i \). The jacknife can be shown to be less biased than the estimator \( \hat{\theta} \). (Smith and Belle 1984).

3. Compute the pseudovalue.

4. Steps 1 to 3 are repeated n times for i=1,...,n. The jacknife estimate \( J_n'(\theta) \) is then given by

\[
J_n'(\theta) = \frac{1}{n} \sum \hat{\theta}_i
\]
However, the standard error of \( \hat{\theta} \) is most explicitly given, on
the other hand, the jacknife method can be used for the
estimation of the standard error of the appropriate estimator.

The jacknife method divides the data of \( M \) observations
into (ideally \( y = M \)) sub groups and obtains \( y' \) partial
estimates by omitting one subgroup at a time. Then pseudo­
values are computed from the resulting samples and the
jacknife estimator is the mean of pseudo-values (Paul
1982).

The jacknife method has successfully been applied using
the Fourier series model for analyzing line transact data
(Buekland, 1982). He found that the jacknife method
performs well when there is a reasonably large sample of
randomly positioned transects. The application of this
method has also been well used by Burnham et al. (1980) in
an application of the jacknife method in the field of
wildlife science.

Thus, over the past two decades considerable research
has been devoted to studying the properties of the jacknife
techniques which was introduced by Quenouille (1949) as a
method of bias reduction and which was later proposed as a
method for robust interval estimations by Tukey. Conditions
under which the jacknife estimator is asymptotically
normally distributed with a consistent estimate of its
variance have been established. It's performance in the
This estimate is known as the first order jacknife and is useful for reducing bias of order $1/n$.

The estimate of the sampling variance of this estimate is given by

$$\text{Var est}(J_n(\theta)) = \frac{1}{n} \sum_{i=1}^{n} (J_n(\theta) - \hat{\theta}_i)^2 / \{n(n-1)\}$$

Chucany, Gray and Owen (1971) generalized the jacknife to remove higher order bias by first removing one element at a time, then removing $n/2$ groups of size 2 and so forth. If this is done for groups of size $k$, the bias is removed to order $O(1/n^k)$. The usual generalized formula is

$$J_n^k(\theta) = (k!)^{-1} \sum_{j=0}^{k} \frac{\hat{\theta}}{\hat{\theta}} (i)(-1)^{j} \binom{k}{j} (n-j)^{k}$$

where $\hat{\theta}$ is the mean of the estimates based on removing groups of size $j$.

Thus, the jacknife technique can reduce the bias of certain estimators and provide significance tests and approximate confidence intervals. The original motivation for the use of the jacknife was to reduce the bias of an estimator and to produce an estimated standard error for the estimator. Miller (1974), has given the general review of literature on jacknife along with some speculations and suggestions for future work.

A natural estimator used for the proportion is $\hat{\rho} = X/n$, where $X = \sum_j X_j$ and $n = \sum_j n_j$. This estimator is unbiased, so the first motivation is unimportant in this context.
problems of ratio estimation and comparison of variances have been studied at length. The jacknife has since been extended to handle estimation in specialized stochastic processes (Gray 1972).

There are many types of jacknifing depending on the nature of the problem. The usual and the most often encountered ones are the unbalanced jacknifing (Miller 1974a), second order jacknifing, generalized jacknifing, and the multisample jacknifing (Miller 1974b).

Some of the common characteristics of all types of jacknifing are that:

(a) Jacknife tests are only asymptotically distribution free.

(b) The jacknife tests are generally more efficient than the Moses test and less efficient than the bootstrap for large samples.

(c) Neither the Moses test nor bootstrap nor jacknife are desirable for small samples (Miller, 1974b, Efron 1981)

Comparisons

Parr (1983) compared the jacknife, bootstrap (when the sample is small and large) and normal theory to the true standard error of for the estimated means of estimators of standard error of sample correlation along with the estimated standard error estimators. Accordingly, th
normal theory method was found superior. Neither the jackknife nor the bootstrap is clearly superior to the other, but certain clear patterns have emerged. The jackknife gave the most variable estimators followed by the bootstrap when the sample size is small followed by the large sample size. The differences were not significant between the jackknife and the small sample size bootstrapping. He also found that at this stage (i.e. when the sample size for bootstrapping is small) computational time was more or less the same. Agreements between the jackknife, bootstrap and the normal theory standard error estimators was quite good especially with correlations greater than 0.8. The sample correlation coefficient ($\gamma$) was examined for sample sizes 14 and 20 consecutively from bivariate normal distributions with population correlations of 0, 0.5 and 0.9.

Efron (1981) discussed the jackknife, the bootstrap and other methods (Bootstrap $N = 128$ and 512, normal smoothed bootstrap, uniform smoothed bootstrap with $N = 128$ and 512, half samples, random half samples, balanced half samples, complimentary half samples, random subsampling and the normal theory). These different techniques have been further defined and have been illustrated with examples. The subject of jackknife (jackknife and infinite jackknife) have been reviewed and summarized by Miller (1974). In this paper more ambitious nonparametric accuracy statements, such as confidence intervals are mentioned without any theory.
In general, the following conclusions have been achieved by the author. The bootstrap performed best among the non-parametric methods aforementioned when explained in terms of the standard error of the estimate (for detailed summary of the comparison between these methods Table 1 of the original paper of Efron 1981 should be referred.)

Efron (1982) compared four different methods (the bootstrap, jackknife, Delta method, normal theory) to the true value on the expected value, standard deviation and the square root of the mean square error (Table 1).

TABLE 1 Comparison of Bootstrap, Jackknife, Delta Method and Normal Theory to the True Value

<table>
<thead>
<tr>
<th></th>
<th>Exp value</th>
<th>Std. Dev.</th>
<th>$\sqrt{MSE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bootstrap</td>
<td>.206</td>
<td>.063</td>
<td>.064</td>
</tr>
<tr>
<td>Jacknife</td>
<td>.223</td>
<td>.085</td>
<td>.085</td>
</tr>
<tr>
<td>Delta Method</td>
<td>.175</td>
<td>.058</td>
<td>.072</td>
</tr>
<tr>
<td>Normal Theory</td>
<td>.247</td>
<td>.056</td>
<td>.056</td>
</tr>
<tr>
<td>True Value J(F)</td>
<td>.218</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

He concluded that among these resampling methods the bootstrap performed the best by having lower MSE. Further illustration has been given in this paper using an example of a medical related problem in discriminant analysis.

Efron (1982) compared the bootstrap and jackknife based on a large Monte Carlo Study. Two samples sizes ($n = 10, 20$) and three distributions ($F \sim N(0, 1), G_1, e^N(0,1)$ were
investigated. Moreover $B = 200$ bootstrap replicates were taken from each trial. In his investigations the bootstrap out performed jackknife except for the small sample size $(n = 10)$ where both of them didn't perform very well. The following table is a summary of his investigation.

TABLE 2. Estimates of Standard Deviation for Switzer's Adaptive Trimmed Mean using the Jacknife and the Bootstrap. The Minimum Possible Coefficient of Variation for a Scale Invariant Estimate of Standard Deviation, Assuming full Knowledge of the Parametric Family, is Shown for $F \sim N(0, 1)$ and $G_1$.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>$n = 10$</th>
<th></th>
<th>Sample Size</th>
<th>$n = 20$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ave</td>
<td>Std. dev.</td>
<td>Ave</td>
<td>Std. dev.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[Coeff var]</td>
<td></td>
<td></td>
<td>[Coeff var]</td>
</tr>
<tr>
<td>F. $(0,1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jacknife</td>
<td>0.327</td>
<td>0.127</td>
<td>0.236</td>
<td>0.070</td>
<td>0.204</td>
</tr>
<tr>
<td>SD</td>
<td>0.296</td>
<td>0.173</td>
<td>0.234</td>
<td>0.143</td>
<td>0.228</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>0.541</td>
<td>0.310</td>
<td>0.339</td>
<td>0.047</td>
<td>0.142</td>
</tr>
<tr>
<td>B=200</td>
<td>0.483</td>
<td>0.310</td>
<td>0.317</td>
<td>0.072</td>
<td>0.317</td>
</tr>
<tr>
<td>True Sd</td>
<td>0.336</td>
<td>0.306</td>
<td>0.224</td>
<td>0.222</td>
<td>0.317</td>
</tr>
<tr>
<td>[Min. possible Cv]</td>
<td>0.24</td>
<td>0.33</td>
<td>0.16</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>No. of trials</td>
<td>1000</td>
<td>3000</td>
<td>1000</td>
<td>3000</td>
<td>1000</td>
</tr>
</tbody>
</table>

Efron (1982) discusses this relationship between the jacknife and bootstrap estimates of the standard deviation and concludes first that jacknife requires less computation than the bootstrap because it approximates any $\hat{e}$ with a linear functional. Secondly, $\hat{\text{SD}}$ jack $\hat{e}$ is itself almost a bootstrap SD estimate, equaling

$$\left[\frac{n}{n-1}\right]^{1/2} + \text{SDJack} (\text{Lin})$$

where the factor makes
factor \( \left[ \frac{n}{n-1} \right]^{1/2} \) makes \( \hat{\Theta} \) unbiased for \( \hat{\Theta}^2 \) if \( \hat{\Theta} \) is a linear function. Proof of the statement has been given in the original paper.

**Bootstrap in Relation to Median**

Efron (1979, and 1982) discusses the theoretical calculation of the bootstrap theory for the median in such a way that it can be calculated without recourse to Monte Carlo methods. It is convenient to consider odd sample sizes say \( n = (2m-1) \) then the sample median \( \hat{\Theta} \) equals \( X_m \), the middle order statistic.

Let \( X(1), X(2), \ldots, X(n) \) be the ordered statistic of an identically and independently distributed sample with distribution \( g(x) \) and

Let also

\[ b_k, n(p) = \binom{n}{k} p^k (1-p)^{n-k} \]

and the random variable \( Z = \# (X_i < e) \) has a binomial distribution with \( p = .5 \), \( Z \sim B_i(n,1/2) \), therefore

\[ \text{Prob}(X(k) < \theta \leq DC(k2)) = \sum_{k=k_1}^{k_2} b_k, n(.5). \]

Since the event \( \{X(k) < \theta \leq X(k)\} \) is the same event \( \{k_1 \leq Z < k_2\} \)

Applied to the median the bootstrap sample \( X_1^*, X_2^*, \ldots, X_n^* \) iid \( F \) has bootstrap sample median \( \hat{\Theta}^* = X^*_m \), the nth ordered value of the \( X_i^* \) even if there are ties among the
Define:
\[ M_j^* = \# \{ X_i^* = X(j) \}, \text{ where } j = 1, 2 \ldots n. \]
The event
\[ \{ X_m^* > X(k) \} \]
is equivalent to
\[ \sum_{j=1}^{k} M_j \leq m - 1, \]
so that
\[ \text{prob} \{ \theta^* > X(k) \} = \text{prob} \sum_{j=1}^{k} \{ M_j < m - 1 \} \]
\[ = \text{prob} \{ B_i(n, k/n) \leq m - 1 \} \]
\[ = \sum_{j=0}^{m-1} \Sigma b_{j, n} (k/n) \]

Therefore the bootstrap distribution of \( \theta^* \) is concentrated on the values \( X(1) < X(2) < \ldots < X(n) \) with bootstrap probability, say, \( p(k) \) of being equal to \( X(k) \).

Then
\[ p(k) = \text{prob} \{ \theta^* = X(k) \} = \sum_{j=0}^{m-1} \binom{k-1}{n-j, n} - \binom{k}{n} \]

He has further illustrated with an example for \( n = 13 \) the bootstrap distribution being as follows:

\[
\begin{array}{cccccccccccc}
(K) & .0000 & .0015 & .0142 & .0550 & .1242 & .1936 & .2230 & .1936 & .1242 & .0550 & \text{(K)} \\
K & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & \text{K} \\
11 & 12 & 13 & 0.0142 & .0015 & .000
\end{array}
\]
Thus, the bootstrap estimate of standard deviation is

\[ \hat{\sigma}_{\text{Boot}} = \left[ \text{P}(k)X^2(k) - (\text{P}(k)X(k))^2 \right]^{1/2} \]

and this can be proved to be asymptotically consistent for the true standard deviation of \( \hat{\sigma} \), in contrast to the jackknife SD estimate:

**MEDIAN FOR THE JACKNIFE**

Estimation of the variance of the sample median is a problem that may arise in several kinds of situations. While it is possible to set a confidence interval for the population median without knowing the variance of the sample median, it may, nevertheless, be useful to report an estimate of this variance. Maritz (1974) has given the following suggestion for estimating the variance of a jackknife estimate.

If the continuous random variable \( X \) has a density of \( f(x) \) and median \( \theta \) then large sample variance of \( X(n) \), the sample median of \( n \) observations is known to be approximately

\[ \frac{1}{4nf^2(\theta)} \]

case where \( X \) is distributed uniformly on the interval \((-1/2, 1/2)\) and \( n = 3 \), the exact variance of \( X(2) \) is 1/20 while the large sample approximation is 1/12. Maritz and Jarrett (1978) have indicated the mathematical approaches to the
procedures to be followed when the sample is odd and even (Maritz 1974).

Efron (1982) shows that Jacknife estimate of the median estimator is

$$\text{Var}(\hat{\theta}_i - 1) = \frac{1}{4f^2(\theta)} \left[ \frac{X_2^2}{2} \right]^2$$

where $X_2^2$ is a Chi deferred in with with 2 df and $E(\frac{X_2^2}{2})^2 = 2$. Therefore the Jacknife estimator is not consistent.

The bootstrap estimator performs reasonably well.

Confidence Interval

The method for evaluating the jacknife confidence interval is discussed in this section.

The jacknife technique may be considered as a nonparametric procedure which can reduce the bias of certain estimators and provide significance tests and approximate confidence intervals (Gary and Schucan, 1972). The classical normal theory F-test for equality of variance is not robust with respect to the assumption of normality. When the underlying populations are not normal the true level of an F-test that is supposed to be of size $\alpha$ may be quite far from $\alpha$. Furthermore, there exists non-normal populations in which, even with large samples, the level of the F-test will not be what it is supposed to be.

The jacknife process for evaluating the lower and upper bound of a confidence interval for variance is based on the
following idea. The pseudo values used in the estimation are also used to derive the C.I. Let $Y_1, Y_2, \ldots, Y_n$ be a sample of independent and identically distributed random variables. Let $S$ be an estimator of the parameter $\theta$ based on this sample. Let $S_i$ be the corresponding estimator based on the sample of size $N-1$ obtained by omitting $Y_i$.

Define:

$$S_i = \frac{\hat{\theta}}{N} - (N-1)S_i$$  

(the pseudo value)

Then:

$$\frac{\hat{\theta}}{S} = \frac{\sum S_i}{N}$$ is the jacknife estimator of $\theta$.

Gary and Schucany (1972) gave the following procedure for jacknife unbiased confidence interval of variance.

Define:

$$S^2 = \frac{N}{N-1} \left( \sum (Y_i - \bar{Y})^2 \right)$$

$$\overline{Y} = \frac{\sum Y_i}{N}$$

$$S_i = \frac{N}{N-1} \left( \sum_{j \neq i} (Y_i - \bar{Y})^2 \right)$$

$$\overline{Y_i} = \frac{\sum Y_i}{N}$$

$$S_i = \frac{N}{N-1} \left( \sum_{j \neq i} (Y_i - \bar{Y})^2 \right)$$

$$\overline{Y_i} = \frac{\sum Y_i}{N}$$

$$\hat{S}_i = N^2 - (N-1)S_i$$

$$V = \frac{\sum (\hat{S}_i - \hat{S})^2}{N-1}, \quad \hat{S} = \frac{\sum S_i}{N}$$

The C.I. is then given by:

$$\left( \hat{S} - t(\gamma, N)V, \hat{S} + t(\gamma, N)V \right)$$
Where that \( t(\gamma,N) \) is the upper percentile of the t-student distribution. Then the interval confidence coefficient approximately \( 1 - \gamma \) will cover the variance with approximately \( 1 - \gamma \) confidence coefficient. Wolfe and Hollander (1973) gave the following modification procedure for a jackknife confidence interval. Suppose we have a sample of \( N \) independent observations, then a positive integer should be selected such that \( N \) is an integer (Choose \( N=k \) in case \( N \) is odd) and divide the observations into \( n \) groups of size \( k \).

\[
S_i^2 = \ln \frac{\sum_{s=1}^{d} (X_{is} - \bar{X}_i)^2}{(d-1)}
\]

where \( \bar{X}_i = \frac{\sum_{s=1}^{d} X_{is}}{N} \)

Define:

\[
S^2 = \ln \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N-1}
\]

Where:

\[
\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}
\]

Compute:

\[
A_i = MS_0^2 - (m-1)S_i^2
\]

Set:

\[
\bar{A} = \frac{\sum_{i=1}^{M} A_i}{L}
\]
and \( V = \frac{\sum_{i=1}^{M} (A_i - \bar{A})^2}{M(M-1)} \)

it can be shown that (independent of the distribution of population)

\[ A \]

\[ \frac{\bar{A}}{V^{1/2}} \]

is approximately normally-distributed with mean zero and variance of one. When \( K = 1 \), \( S_i \) is the natural log of the \((N-1)\) X observations remaining when \( X_i \) is deleted for the total \( X \) sample that is \( S_i \) is the natural log of the sample variance for the observations \( x_1, \ldots, x_i - 1, x_i + 1, \ldots x_n \)

then the approximate confidence interval for \( \sigma^2 \) with confidence coefficient \( 1 - \gamma \) based on the jacknifed procedure is given by

where \( \sigma^2_L = \exp (\bar{A} - Z_{\alpha/2} V^{1/2}) \)

\( \sigma^2_u = \exp (\bar{A} + Z_{\alpha/2} V^{1/2}) \)

Wolfe and Hollander (1973) believe that replacing \( Z_{\alpha/2} \) by \( t(n,\alpha) \) for small number of data will improve the accuracy of confidence interval distribution. The natural logs were used basically to reduce and stabilize the variance.
APPLICATIONS

Both of these resampling techniques are currently and widely used in ecological studies, wildlife, Veterinary sciences, Astrophysics, Range sciences, and many other natural and physical sciences. Specifically, the advent of computers and statistical computations has recently advanced the use of bootstrapping and it has quite recently got much of the literature attention and started being widely used for research purposes.

CONCLUSION

Two resampling techniques, bootstrap and jackknife, are compared on the bases of their bias, estimation of the variance of the sample median and placement of confidence intervals. Moreover, procedural approaches have also been discussed.

Three kinds of bootstrapping (ordinary, randomized, and double bootstrapping techniques have been elucidated. Bootstrapping and jacknifing work best with very large samples but usually even larger samples are required for bootstrap. Biases, and MSE associated with bootstrapping are lower than the jackknife. The estimates of variance of the sample median works best with bootstrapping. In addition, the bootstrapping has narrower confidence intervals. The only demerit associated with all kinds of bootstrapping is the higher computational time required compared to the jackknife.
REFERENCES


APPENDIXES
TYPE B: BOOT.EX

MTB > READ C1 C2
DATA> 3.2
DATA> 5.2
DATA> 8.2
DATA> 11.2
DATA> 2.2
DATA> PRINT C1-C2
5 ROWS READ
ROW C1 C2
1 3 0.2
2 5 0.2
3 8 0.2
4 11 0.2
5 2 0.2

MTB > HELP DRAN
DRANDOM K obs. using values in C and prob. in C, put into C

Simulates one sample of K observations from a discrete distribution
specified by the two columns. You must input your chosen probability
function before using this command.

MTB > DRAN 5 C1 C2 C3
5 DISCRETE RANDOM OBSERVATIONS

MTB > DRAN 5 C1 C2 C4
5 DISCRETE RANDOM OBSERVATIONS

MTB > DRAN 5 C1 C2 C5
5 DISCRETE RANDOM OBSERVATIONS

MTB > PRINT C1-C5
ROW C1 C2 C3 C4 C5
1 3 0.2 11 5 2
2 5 0.2 2 5 3
3 8 0.2 2 5 3
4 11 0.2 8 2 2
5 2 0.2 8 8 2

MTB > DESC C1-C5

MTB > AVER C3 K3
MEAN = 5.2000

MTB > AVER C4 K4
MEAN = 4.6000

MTB > AVER C5 K5
MEAN = 5.2000

MTB > JOIN K3 K4 K5 C6

MTB > PRINT C6
C6
5.2 4.6 6.2

MTB > AVER C6
MEAN = 5.3333
ST.DEV. = 0.80829