A Comparison of Bootstrap Methods for Variance Estimation

Saeid Amiri, Dietrich von Rosen, and Silvelyn Zwanzig

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2009:02
ISSN 1651-8543

A Comparison of Bootstrap Methods for Variance Estimation

Saeid Amiri[1], Dietrich von Rosen

Centre of Biostochastics
Swedish University of Agricultural Sciences
P.O. Box 7032, SE-750 07 Uppsala, Sweden

Silvelyn Zwanzig

Department of Mathematics
Uppsala University, P.O.Box 480, 751 06 Uppsala, Sweden

Abstract

This paper presents a comparison of the nonparametric and parametric bootstrap methods, when the statistic of interest is the sample variance estimator. Conditions when the nonparametric bootstrap method of variance performs better than the parametric bootstrap method are described.

Keywords: Bootstrap, nonparametric, parametric, kurtosis.

[1]E-mail address to the correspondence author: saeid.amiri@et.slu.se

1 Introduction

There has been much theoretical and empirical research on properties of the bootstrap method and it has become a standard tool in statistical analysis. The idea behind the bootstrap method is that if the sample distribution is a good approximation of the population distribution, the sampling distribution of interest can be used to generate a large number of new samples from the original sample via sampling with replacement. Bootstrapping treats the sample as the actual population.

The most important property of the bootstrap method is the ability to estimate the standard error of any well-defined function of the random variables corresponding to the sample data. Applying the bootstrap method requires fewer assumptions than are needed for conventional methods. There are many books and papers on the bootstrap method and its applications in a variety of fields, see e.g. Hall (1992), Efron and Tibshirani (1993), Shao and Tu (1996), Davison and Hinkley (1997), Mackinnon (2002), Janssen and Pauls (2003) and Athreya and Lahiri (2006). Many use the bootstrap method without focusing on the theory. However by considering theoretical aspects, it is possible to understand the mechanism behind the simulations.

In this paper we present a finding that helps explain the different in performance of the nonparametric and parametric bootstrap methods. The statistic of interest is the variance. The commonly applied nonparametric bootstrap resamples the observations from an original sample, whereas the parametric bootstrap method generates bootstrap observations by a given parametric distribution. If justification cannot be provided for the use of a specific parametric distribution, then a nonparametric bootstrap can be used. This is discussed in the rest of this paper.

The nonparametric and parametric methods are simultaneously considered in some studies, but often the results of the simulations are given without explicit discussions of their different performances. For example, Efron and Tibshirani (1993) discuss the nonparametric and parametric bootstrap confidence intervals of variance by using an example, Ostaszewski and Rempala (2000) explain how to use the bootstrap methods within the actuary sciences and Lee (1994) explains how to use a tuning parameter to find more accurate estimation.

It is difficult to study the comparison in general but it is possible for the main parameters such as the mean and variance. Here, We will use a heuristic criterion to compare the bootstrap method with the real distribution. Although the bootstrap method is based on the sample, it is intended to approach the real distribution. According to Hall (1992), the bootstrap method may be expressed as an expectation conditional on the sample or as an integral with respect to the sample distribution function. This allows us to make direct comparisons of the nonparametric and parametric bootstrap method and to draw conclusions from these comparisons.

We can show that the behavior of the nonparametric and parametric bootstrap methods of variance estimation is affected by the kurtosis, which is explained in Theorem 1. This can be expected because the variance of variance depends on the fourth moment. Distributions are usually classified by how flat-topped the distribution is relative to the normal distribution. This can be done via the sample kurtosis. It should be mentioned that there is not a universal agreement about what the kurtosis is, see Darlington (1970) and Joanes and Gill (1998). In the case of variance estimation, we show that the bootstrap estimation depends on the kurtosis. However there is no difference between the parametric and nonparametric bootstrap method in the case of mean estimation. In the case of variance estimation, we show that the nonparametric bootstrap method can be better than the parametric bootstrap under some conditions, regardless whether the real distribution and the distribution of the parametric bootstrap method belong to the same distribution family.

In Section 2, we briefly outline the bootstrap approaches. In Section 3, the main results are presented. In Section 4, the theoretical discussion is illustrated using some examples.

2 Bootstrap method

Let us look at the bootstrap stages, which can be formulated as below:

1. Suppose $\mathcal{X} = (X_1, \cdots, X_n)$ is an i.i.d. random sample of the distribution F. Assume $V(X) = \sigma^2$ and $EX^4 < \infty$.

2. We are interested in $\theta(F) = \sigma^2$ and consider plug-in estimation: $\widehat{\theta} = \theta(X_1, \cdots, X_n) = \theta(F_n) = S_X^2$,
$$S_X^2 = \frac{1}{n}\sum_{j=1}^{n} X_j^2 - (\overline{X})^2, \tag{1}$$
where F_n is the empirical distribution function, i.e. $F_n(x) = \frac{1}{n}\sum_{j=1}^{n} I(X_j \leq x)$.

3. Generate the bootstrap samples. This can be done in two different ways, the nonparametric and parametric bootstrap, with the symbols "$*$" and "#" used to distinguish the approaches.

(i) The nonparametric bootstrap method: $X^*_{ij} \overset{iid}{\sim} F_n(x), i = 1, \ldots, B$, $j = 1, \ldots, n$. Note that if $Z \sim F_n(x)$ then $EZ = \overline{X}$ and $V(Z) = S^2_X$, where S^2_X is the second centered moment estimator. The kurtosis of F_n is defined as:

$$K_{F_n} = \frac{\sum_{j=1}^{n}(X_j - \overline{X})^4/n}{\left(\sum_{j=1}^{n}(X_j - \overline{X})^2/n\right)^2}. \tag{2}$$

(ii) The parametric bootstrap method: $X^{\#}_{ij} \overset{iid}{\sim} G_{\widehat{\lambda}}, i = 1, \ldots, B,\ j = 1, \ldots, n$ where $G_{\widehat{\lambda}} = G(.|\mathcal{X})$ is an element of a class $\{G_\lambda, \lambda \in \Lambda\}$ of distributions. The parameter λ is estimated by statistical methods. We also have $E(X^{\#}) = \overline{X}$ and $V(X^{\#}) = S^2_X$. In this case, the kurtosis, $K_{G(.|\mathcal{X})}$, is defined as:

$$K_{G(.|\mathcal{X})} = \frac{E_{\mathcal{X}}(X - \overline{X})^4}{(E_{\mathcal{X}}(X - \overline{X})^2)^2},$$

where $E_{\mathcal{X}}(.) = E(.|\mathcal{X})$ is the conditional expectation.

4. Calculate the bootstrap replications

$$S^2(X^\times_i) = S^2(X^\times_{i1}, \ldots, X^\times_{in}) \qquad i = 1, \ldots, B,$$

The symbol $\times$ is used when either the parametric or nonparametric procedure hold.

5. Handle the bootstrap replications as i.i.d. random samples and consider the sample mean and the sample variance. They are:

$$S^{2\times} = \frac{1}{B}\sum_{i=1}^{B} S^2(X^\times_i), \tag{3}$$

$$V^\times = \frac{1}{B}\sum_{i=1}^{B}\left(S^2(X^\times_i) - S^{2\times}\right)^2 = \frac{\sum_{i=1}^{B}\left(S^2(X^\times_i)\right)^2}{B} - (S^{2\times})^2. \tag{4}$$

It is obvious that $V^\times$ measures the bootstrap variation which shows how the replications settle around the mean. The square root of $V^\times$ is referred to as the bootstrap estimate of standard error, se_B.

It is difficult to offer a specific guideline to compare the variances explicitly. Here e is a proposed criterion in order to compare the ratio of the conditional expectation of V^* and $V^{\#}$, i.e.

$$e = \frac{E(V^*|\mathcal{X})}{E(V^{\#}|\mathcal{X})}. \tag{5}$$

Thus it is suggested to compare the expectation of the variances of sample variances. When $e < 1$, it means that the replications of the nonparametric bootstrap concentrate more than the replications of the parametric bootstrap.

3 Comparison of the bootstrap methods

Two statistics that are often used to study properties of estimators are biasedness and mean square error (MSE) which are studied here for the bootstrap estimation approaches of the variance. Let us first look at biasedness.

3.1 Biasedness

The following theorem clarifies the properties of the nonparametric and parametric bootstrap estimators of variance. It can be shown explicitly how bootstrapping is affected by the kurtosis.

Theorem 1 *Let* $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F$ *with* $EX^4 < \infty$. *Then for the bootstrap methods in (i) and (ii), presented in the previous section:*

$$E(S^{2*}|\mathcal{X}) \quad = \quad E(S^{2\#}|\mathcal{X}) = \frac{n-1}{n} S^2_X, \tag{6}$$

$$K_{F_n} < K_{G(.|\mathcal{X})} \quad \Longleftrightarrow \quad E(V^*|\mathcal{X}) < E(V^{\#}|\mathcal{X}), \tag{7}$$

where K_{F_n} *and* $K_{G(.|\mathcal{X})}$ *are the sample kurtosis and the kurtosis corresponding to the parametric distribution* $G_{\hat{\lambda}}$ *used in (ii).*

Proof: By the construction in (i) and (ii), the conditional expectation of $S^2(X_i^\times)$ and $S^{2\times}$ given in (3) are as follow:

$$\begin{aligned} E\big(S^2(X_i^\times)|\mathcal{X}\big) &= E_{\mathcal{X}}\Big(\frac{1}{n}\sum_{j=1}^{n} X_{ij}^{\times 2} - (\overline{X}_i^\times)^2\Big) \\ &= \frac{n-1}{n}\Big(E_{\mathcal{X}}(X^{\times 2}) - \big(E_{\mathcal{X}}(X^\times)\big)^2\Big), \end{aligned}$$

and according to (3)

$$E_{\mathcal{X}}(S^{2\times}) = E_{\mathcal{X}}\big(\frac{1}{B}\sum_{i=1}^{B} S^2(X_i^\times)\big) = E_{\mathcal{X}}\big(S^2(X_i^\times)\big). \tag{8}$$

Therefore

$$E(S^{2*}|\mathcal{X}) = E(S^{2\#}|\mathcal{X}) = \frac{n-1}{n}S_X^2,$$

and (7) is verified. The conditional expectation of $V^\times$ is given by:

$$\begin{aligned}
E_{\mathcal{X}}(V^\times) &= \frac{1}{B}\sum_{i=1}^{B} E_{\mathcal{X}}(S^2(X_i^\times)^2) - E_{\mathcal{X}}((S^{2\times})^2) \\
&= E_{\mathcal{X}}(S^2(X_i^\times)^2) - \big[\frac{1}{B}E_{\mathcal{X}}(S^2(X_i^\times)^2) + \frac{B-1}{B}E_{\mathcal{X}}((S^{2\times})^2\big] \\
&= \frac{B-1}{B}\big[E_{\mathcal{X}}(S^2(X_i^\times)^2) - \big(E_{\mathcal{X}}(S^{2\times})\big)^2\big], \qquad (9)
\end{aligned}$$

where

$$\begin{aligned}
&E_{\mathcal{X}}(S^2(X_i^\times)^2) \\
&= E_{\mathcal{X}}\big(\frac{1}{n^2}\big[(\sum_{j=1}^{n} X_j^{\times 2})^2 + n^2\overline{X}^{\times 4} - 2n\overline{X}^{\times 2}\sum_{j=1}^{n} X_j^{\times 2}]\big) \\
&= \frac{1}{n^4}\big[(n-1)^2 n E_{\mathcal{X}}(X^{\times 4}) + (4-4n)n(n-1)E_{\mathcal{X}}(X^{\times 3})E_{\mathcal{X}}(X^\times) \\
&\quad + n(n-1)(n^2+3-2n)(E_{\mathcal{X}}(X^{\times 2}))^2 \\
&\quad + 3(12-4n)\begin{pmatrix} n \\ 3 \end{pmatrix} E_{\mathcal{X}}(X^{\times 2})(E_{\mathcal{X}}(X^\times))^2 + 24\begin{pmatrix} n \\ 4 \end{pmatrix}(E_{\mathcal{X}}(X^\times))^4\big] \\
&= \big(\frac{n-1}{n^3}\big)\big[(n-1)\big(E_{\mathcal{X}}\big(X^\times - E_{\mathcal{X}}(X^\times)\big)^4\big) \\
&\quad + (n^2-2n+3)\big(E_{\mathcal{X}}(X^\times - E_{\mathcal{X}}(X^\times))^2\big)^2\big]. \qquad (10)
\end{aligned}$$

By using (8) and (10),

$$\begin{aligned}
E_{\mathcal{X}}(V^\times) &= \frac{B-1}{B}\big(E_{\mathcal{X}}(S^2(X_i^\times)^2) - (E_{\mathcal{X}}(S^{2\times}))^2\big) \\
&= \big(\frac{B-1}{B}\big)\big(\frac{n-1}{n^3}\big)\big[(n-1)\big(E_{\mathcal{X}}\big(X^\times - E_{\mathcal{X}}(X^\times)\big)^4\big) \\
&\quad - (n-3)\big(E_{\mathcal{X}}(X^\times - E_{\mathcal{X}}(X^\times))^2\big)^2\big] \\
&= \big(\frac{B-1}{B}\big)\big(\frac{n-1}{n^3}\big)\big(E_{\mathcal{X}}(X^\times - E_{\mathcal{X}}(X^\times))^2\big)^2\big((n-1)K^\times - (n-3)\big) \\
&= \big(\frac{B-1}{B}\big)\big(\frac{n-1}{n^3}\big)\big(S_X^2\big)^2\big((n-1)K^\times - (n-3)\big), \qquad (11)
\end{aligned}$$

where $K^{\times}$ can be either K_{F_n} or $K_{G(.|\mathcal{X})}$. Thus the difference between $E_{\mathcal{X}}(V^*)$ and $E_{\mathcal{X}}(V^{\#})$ depends on the kurtosis. The ratio of the conditional expectation of V^* and $V^{\#}$ by using (11) equals:

$$e = \frac{E(V^*|\mathcal{X})}{E(V^{\#}|\mathcal{X})} = \frac{(S_X^2)^2\big((n-1)K_{F_n} - (n-3)\big)}{\big(S_X^2\big)^2\big((n-1)K_{G(.|\mathcal{X})} - (n-3)\big)}, \qquad (12)$$

which implies that:

$$E(V^*|\mathcal{X}) < E(V^{\#}|\mathcal{X}) \Longleftrightarrow K_{F_n} < K_{G(.|\mathcal{X})}. \qquad (13)$$

□

From relation (7), it follows that the unconditional expectation of the bootstrap parametric and nonparametric estimator equal

$$E(S^{2*}) = E(S^{2\#}) = \left(\frac{n-1}{n}\right)E(S_X^2) = \left(\frac{n-1}{n}\right)^2\sigma^2. \qquad (14)$$

Thus

$$E(S^{\times 2}) < E(S_X^2) < \sigma^2,$$

where σ^2 is the variance of the underlying distribution. Therefore the bootstrap estimator of variance is more biased than the sample variance. The following example clarifies this result in the case of the normal distribution.

Example 1: Consider the parametric bootstrap with the normal distribution, namely $X_{i,j}^{\#}|\mathcal{X} \sim N(\overline{X}, S_X^2)$. Then $\frac{nS^2(X_i^{\#})}{S_X^2}|\mathcal{X} \sim \chi_{n-1}^2$, which implies $E_{\mathcal{X}}\big(S^2(X_i^{\#})\big) = \frac{(n-1)S_X^2}{n}$ and $K_{G(.|\mathcal{X})} = 3$. Therefore in this case, a closed form of the expectation of the parametric bootstrap estimator can be found:

$$E_{\mathcal{X}}\big(S^{2\#}\big) = E_{\mathcal{X}}(\frac{1}{B}\sum_{i=1}^{B} S^2(X_i^{\#})) = \frac{1}{B}\sum_{i=1}^{B} E_{\mathcal{X}}(S^2(X_i^{\#})) = \frac{n-1}{n}S_X^2.$$

Since $K_{G(.|\mathcal{X})} = 3$ and by (11),

$$E_{\mathcal{X}}(V^{\#}) = 2(\frac{n-1}{n^2})(\frac{B-1}{B})S_X^4. \qquad (15)$$

Hence in the case of the normal distribution, if $K_{F_n} < 3$ holds, then $E_{\mathcal{X}}(V^*)$ is less than $E_{\mathcal{X}}(V^{\#})$. Hence the replications of the nonparametric bootstrap

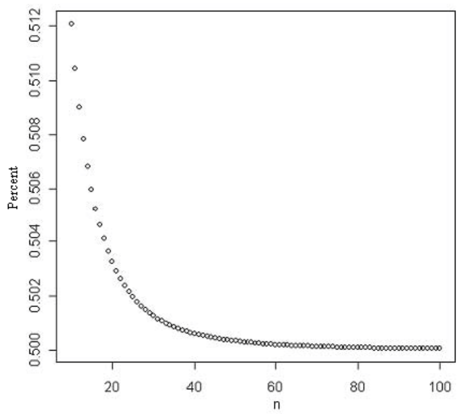

Figure 1: Percentage of times $K_{F_n} < 3$ happen in the simulation as a function of sample size.

method of the variance concentrates more than the replications of parametric bootstrap method.

For the nonparametric and parametric bootstrap method with the normal distribution, the replications of S^{2*} are fairly close in comparison with $S^{2\#}$. This can be explained as follows. Consider the normal distribution $F = N(\mu, \sigma^2)$ and $G(.|\mathcal{X}) = N(\overline{X}, S_X^2)$ where $K_{G(.|\mathcal{X})} = 3$. Cramér (1945) showed that the estimation of the sample kurtosis is negatively biased, i.e.

$$\begin{aligned} E(\widehat{K}_{F_n}) &= 3 - \frac{6}{n+1}, \\ V(\widehat{K}_{F_n}) &= \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}. \end{aligned}$$

Therefore $\widehat{K}_{F_n} < 3$ is "more likely" to occur, which is illustrated in Figure 1. As previously mentioned $\widehat{K}_{F_n} < 3$ makes the nonparametric replications concentrate more than the parametric bootstrap. A discussion of biasedness of the sample kurtosis is difficult. However, by using Monte Carlo simulation it can be shown that $\widehat{K}_{F_n} < K_{G(.|\mathcal{X})}$ is held for the t, chisquare and uniform distributions, as well as the normal distribution.

Theorem 1 provides the possibility to make a direct comparison of the two bootstrap approaches. However one should make a comparison relative to the true underlying distribution. The following theorem helps us to carry out this comparison. In general, the variance of variance is given by:

$$V(S_X^2) = \frac{n-1}{n^3}\sigma^4\big((n-1)K - (n-3)\big), \tag{16}$$

Theorem 2 *Let $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F$ with $EX^4 < \infty$. Then for the bootstrap methods in (i) and (ii) in Section 1 where $K_{G(.|\mathcal{X})}$ is assumed to be independent of observations, the following relations hold for V^* and $V^{\#}$, which are defined in (4):*

$$
\begin{aligned}
E(V^*) &= \frac{B-1}{B}\frac{(n-1)^2}{n^6}\sigma^4\Big(K(n-1)(n^2-4n+6) \\
&\quad + (-n^3+11n^2-24n+18)\Big), && (17)\\
E(V^{\#}) &= \frac{B-1}{B}\frac{(n-1)^2}{n^6}\sigma^4\Big(K(n-1)+(n^2-2n+3)\Big) \\
&\quad \Big((n-1)K_{G(.|\mathcal{X})} - (n-3))\Big), && (18)
\end{aligned}
$$

where K is the kurtosis of F.

Proof: The following relations are presented by Cramér (1945, p.349),

$$
\begin{aligned}
E(S^4) &= \frac{n-1}{n^3}\sigma^4\big((n^2-2n+3)+(n-1)K\big), && (19)\\
E\Big(\sum_{i=1}^{n}(X_i-\overline{X})^4/n\Big) &= \frac{n-1}{n^3}\sigma^4\Big((n^2-3n+3)K+3(2n-3)\Big). && (20)
\end{aligned}
$$

The proof of theorem is completed by inserting these equations into the corresponding terms in (11). □

This theorem states that $E(V^*)$ depends on K, whereas $E(V^{\#})$ depends on K and $K_{G(.|\mathcal{X})}$. It should be noted that if $K_{G(.|\mathcal{X})}$ depends on the observations, for example the lognormal distribution, then it is impossible to present a closed form solution. Hence in this case, study of the performance of the parametric bootstrap is rather difficult. However for the nonparametric bootstrap, (17) always holds. It is obvious that the methods are biased. In the case of the normal distribution, the following corollary is a direct result of Theorem 2.

Corollary 1 *If $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F = N(\mu, \sigma^2)$ and also if $G(.|\mathcal{X}) = N(\bar{X}, S_X^2)$, then the following relations hold:*

$$E(V^*) < E(V^{\#}) < V(S_X^2), \tag{21}$$

$$\frac{Bn^3}{(B-1)(n-1)(n^2-2n+3)}E(V^*) = \frac{Bn^3}{(B-1)(n^2-1)n}E(V^{\#}) = V(S_X^2). \tag{22}$$

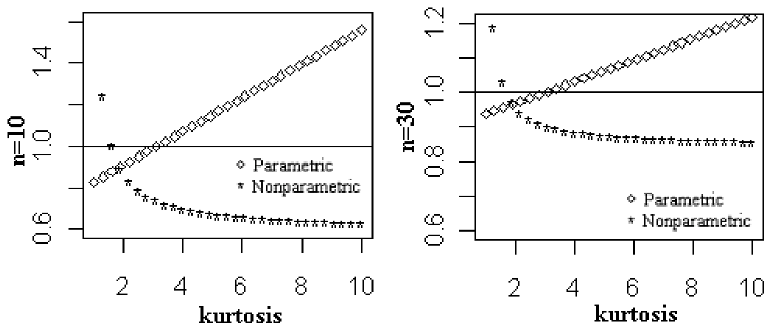

Figure 2: Plots of $\frac{E(V^{\#})}{V(S_X^2)}$ and $\frac{E(V^*)}{V(S_X^2)}$ versus kurtosis, for the sample size of 10 and 30, respectively.

If the underlying distribution of F and $G(.|\mathcal{X})$ belong to the normal distribution family, it is expected that the standard error of the parametric bootstrap of variance will be closer to the variance of F in comparison with the nonparametric bootstrap. Moreover, by using the corrections given in (22), it is possible to find unbiased estimators of the parametric and nonparametric bootstrap of variance.

It is interesting that when $K_{F_n} > 3$ then $E(V^*|\mathcal{X}) > E(V^{\#}|\mathcal{X})$ and since $V(S_X^2)$ is larger than the expected bootstrap estimation, (21), V^* has more chance to be close to $V(S_X^2)$ than $V^{\#}$. This is explained later in this paper by simulations (see Table 4).

Figure 2 clarifies the behavior of $E(V^{\#})$ and $E(V^*)$, which are given in (17) and (18). Plots include $\frac{E(V^{\#})}{V(S_X^2)}$ and $\frac{E(V^*)}{V(S_X^2)}$ for sample sizes of 10 and 30, where the F and $G(.|\mathcal{X})$ belong to the same distribution family. It is obvious that the nonparametric bootstrap method has a negative bias for large sample size against the parametric bootstrap method. The most important result is that for the distribution with the kurtosis approximately $1.4 \leq \widehat{K} \leq 2$, the nonparametric bootstrap method is less biased than the parametric bootstrap method, regardless whether F and $G(.|\mathcal{X})$ have the same distribution. This range is based on n, which can be found by (17) and (18). It is obvious that the amount of kurtosis affects the accuracy of variance estimation.

Now let us study the case when F and $G(.|\mathcal{X})$ do not belong to same distribution family. In Figure 3, $G(.|\mathcal{X}) \sim N(\mu, \sigma^2)$ and F is an arbitrary distribution. It includes $\frac{E(V^{\#})}{V(S_X^2)}$ and $\frac{E(V^*)}{V(S_X^2)}$ for sample sizes of 10 and 30. It clarifies when there is uncertainty about the real distribution, special care

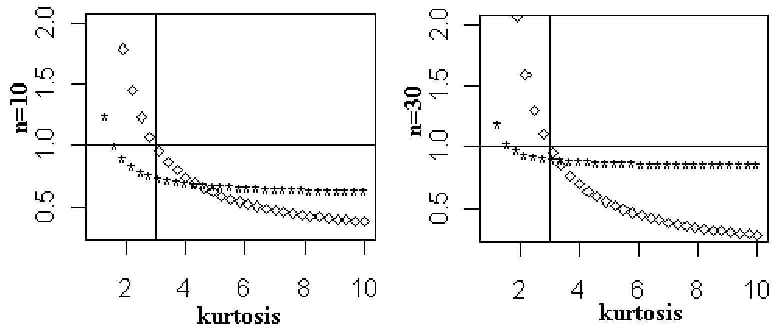

Figure 3: Plot of $\frac{E(V^{\#})}{V(S_X^2)}$ and $\frac{E(V^*)}{V(S_X^2)}$ versus kurtosis, for sample size of 10 and 30, where $G(.|\mathcal{X}) \sim N(\overline{X}, S_X^2)$.

should be taken when using the parametric bootstrap method. It is obvious that the performance of the nonparametric bootstrap method quickly improves with an increase in sample size.

The following example is given to explain why the nonparametric may be better, regardless if $G(.|\mathcal{X})$ and F belong to the same distribution family.

Example 2: The probability density function (pdf) of the exponential power distribution (EPD) family is given as follows (Chiodi, 1995):

$$f_X(x) = \frac{1}{2p^{1/p}\Gamma(1+1/p)\sigma_p} exp(\frac{-|x-\mu|^p}{p\sigma_p^p}) \qquad p > 0,\ x\epsilon R, \tag{23}$$

where μ and σ_p are location and scale parameters and p is a shape parameter. The following relations hold:

$$\mu = E(X), \qquad \sigma_p = E(|X-\mu|^p)^{1/p}, \qquad K = \frac{\Gamma(1/p)\Gamma(5/p)}{\big(\Gamma(3/p)\big)^2}.$$

These can be some known pdfs, e.g. with $p = 1$ it is the Laplace, $p = 2$ it is the normal distribution, and for p$\longrightarrow \infty$ it is the uniform distribution. It is obvious that EPD with $p = 10$ has a kurtosis of 1.884. Let $\mu = 0$ and $\sigma_p = 2$, for this case $\sigma^2 = 2$ and with $n = 30$ then $V(s_X^2) = 0.1187$. Figure 4 is the violin plot of $|V(S^{2\times}) - V(S_X^2)|$ obtained from Monte Carlo simulation of the parametric and nonparametric bootstrap, with $B = 500$ and the Monte Carlo replications based on 1000 simulations. The results show that the nonparametric bootstrap is closer to $V(S^2)$ than the parametric bootstrap.

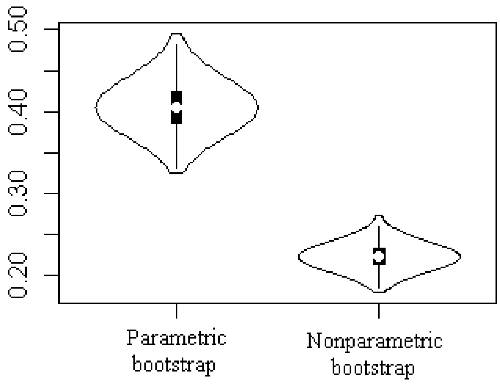

Figure 4: Violin plot of the nonparametric and parametric bootstrap of EPD.

The violin plot is a combination of a box plot and a kernel density plot. It helps to study the results of the simulations. The violin plot is a combination of a box plot and a kernel density plot, see Hintze and Nelson (1998).

3.2 MSE

The variability of estimation can also be assessed by its MSE, defined as

$$MSE(\widehat{\theta}) = V(\widehat{\theta}) + (Bias)^2. \tag{24}$$

The MSE of the sample variance is given as follows,

$$MSE(S_X^2) \quad = \quad V(S_X^2) + \frac{1}{n^2}\sigma^4. \tag{25}$$

In the following, first the conditional MSE, which is the direct result of the bootstrap method, is first discussed. Actually the conditional MSE is the direct result of the bootstrap method and the unconditional is a combination of the bootstrap and the frequentist approaches. The following lemma discusses the bootstrap estimation of S^2.

Lemma 1 *Let* $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F$ *with* $E(X^4) < \infty$. *Then for the bootstrap methods in (i) and (ii) in section 2:*

$$\lim_{B \longrightarrow \infty} MSE(S^{2*}|\mathcal{X}) = \lim_{B \longrightarrow \infty} MSE(S^{2\#}|\mathcal{X}) = (S_X^2 - \sigma^2)^2 \tag{26}$$

PROOF: Because of independency of the conditional $S_i^{2\times}$, the following equation holds,

$$V(S^{2\times}|\mathcal{X}) = \frac{1}{B}V(S_i^{2\times}|\mathcal{X}).$$

For $B \longrightarrow \infty$, This tends to zero and therefore $MSE(S^{2\times}|\mathcal{X})$ converges to the squared biasedness which is given in (26). □

Lemma 2 *If $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F$ with $E(X^4) < \infty$. Then for the bootstrap methods explained in (i) and (ii) in Section 2:*

$$\lim_{B\longrightarrow\infty} MSE(S^{2*}) = \lim_{B\longrightarrow\infty} MSE(S^{2\#}) = \left(\frac{n-1}{n}\right)^2 V(S_X^2) + \left(\frac{1-2n}{n^2}\right)^2 \sigma^4. \tag{27}$$

PROOF: The first term of MSE is obtained as follows:

$$\begin{aligned}
\lim_{B\longrightarrow\infty} V(S^{2\times}) &= \lim_{B\longrightarrow\infty} \big(V(E(S^{2\times}|\mathcal{X})) + E(V(S^{2\times}|\mathcal{X}))\big) \\
&= V(\frac{n-1}{n}S_X^2) + \lim_{B\longrightarrow\infty} E\big(\frac{1}{B}V(S^{2\times}(X_i)|\mathcal{X}\big) \\
&= \left(\frac{n-1}{n}\right)^2 V(S_X^2).
\end{aligned}$$

The second term can be obtained directly from (14). □

Lemma 2 clarifies that the MSE of the parametric and nonparametric bootstrap estimators of variance are asymptotically the same. The discussion of $MSE(V^\times)$ is rather difficult because it is based on higher moments. However it can be studied asymptotically. The following lemmas are required in order to establish the proof for the Theorem 3. First the conditional MSE is considered.

Lemma 3 *Let $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F$ and $E(X^8) < \infty$ then*

$$\lim_{B\longrightarrow\infty} MSE(V^*|\mathcal{X}) = \lim_{B\longrightarrow\infty} \big(E(V^*|\mathcal{X}) - V(S^2)\big)^2, \tag{28}$$

$$\lim_{B\longrightarrow\infty} MSE(V^\#|\mathcal{X}) = \lim_{B\longrightarrow\infty} \big(E(V^\#|\mathcal{X}) - V(S^2)\big)^2, \tag{29}$$

where $E(V^*|\mathcal{X})$ and $E(V^\#|\mathcal{X})$ are given in (11).

PROOF: It holds that:

$$\begin{aligned} V(V^{\times}|\mathcal{X}) &= V\left(\frac{1}{B}\sum_{i=1}^{B}(S^{2\times}(X_i) - S^{2\times})^2|\mathcal{X}\right) \\ &= \frac{B-1}{B^3}((B-1)E(S^{2\times}(X_i) - S^{2\times}|\mathcal{X}) \\ &\quad -(B-3)(E(S^{2\times}(X_i) - S^{2\times}|\mathcal{X})^2) \end{aligned}$$

Therefore $V(V^{\times}|\mathcal{X}) \longrightarrow 0$ as $B \longrightarrow \infty$. This leads to:

$$\lim_{B\longrightarrow\infty} MSE(V^{\times}|\mathcal{X}) = 0 + \lim_{B\longrightarrow\infty}(Bias)^2. \tag{30}$$

□

This lemma shows that the conditional $MSE(V^{\times})$ is affected by the kurtosis via $E(V^{\times}|\mathcal{X})$ which is expected by the discussion of Theorem 1. The following lemma is necessary for the discussion of $MSE(V^{\times})$.

Lemma 4 *Let* $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F$ *and* $E(X^8) < \infty$.

$$V(\widehat{V}(S_X^2)) = \left(\frac{n-1}{n^3}\right)^2\left(n^2V(\widehat{\mu}_4) + n^2V(\widehat{\mu}_2^2)\right) + O(n^{-4}), \tag{31}$$

where $\widehat{V}(S_X^2)$ *is the estimate of (16) and*

$$\begin{aligned} V(\widehat{\mu}_4) &= \frac{\mu_8 - 8\mu_3\mu_5 - \mu_4^2 + 16\mu_2\mu_3^2}{n} + O(n^{-2}), \\ V(\widehat{\mu}_2^2) &= \frac{4\mu_4\mu_2^2 - 4\mu_2^4}{n} + O(n^{-2}), \\ \mu_i &= E(X_i - \mu)^i. \end{aligned} \tag{32}$$

PROOF: Let

$$\widehat{V}(S_X^2) = \frac{n-1}{n^3}((n-1)\widehat{\mu}_4 - (n-3)\widehat{\mu}_2^2),$$

where $\widehat{\mu}_2$ and $\widehat{\mu}_4$ are the estimators of the second and fourth centrad moments. Then it can be shown that

$$\begin{aligned} V(\widehat{V}(S_X^2)) &= \left(\frac{n-1}{n^3}\right)^2\Big((n-1)^2V(\widehat{\mu}_4) + (n-3)^2V(\widehat{\mu}_2^2) \\ &\quad - 2(n-1)(n-3)Cov(\widehat{\mu}_4, \widehat{\mu}_2^2)\Big), \end{aligned}$$

where

$$Cov(\widehat{\mu}_4, \widehat{\mu}_2^2) = \frac{1}{n}V(\widehat{\mu}_4) + 2\frac{n-1}{n^2}Cov\big((X_1-\overline{X})^4, (X_2-\overline{X})^2(X_3-\overline{X})^2\big) + \frac{(n-1)(n-2)}{n^2}Cov\big((X_1-\overline{X})^4, (X_1-\overline{X})^2(X_2-\overline{X})^2\big).$$

Moreover, It can be shown by some algebra that:

$$\begin{aligned} & Cov\big((X_1-\overline{X})^4, (X_1-\overline{X})^2(X_2-\overline{X})^2\big) \\ & \quad = (\mu_6\mu_2 - \mu_4\mu_2^2) + \frac{1}{n}(21\mu_4\mu_2^2 - 7\mu_6\mu_2 - 6\mu_2^4) + O(n^{-2}) \\ & Cov\big((X_1-\overline{X})^4, (X_2-\overline{X})^2(X_3-\overline{X})^2\big) \\ & \quad = \frac{1}{n^2}(23\mu_2^2\mu_4 - 85\mu_2^4 + 2\mu_6\mu_2) + O(n^{-3}) \end{aligned}$$

Therefore by using these relations the following result is obtained

$$V(\widehat{V}(S_X^2)) = \left(\frac{n-1}{n^3}\right)^2 \left(n^2V(\widehat{\mu}_4) + n^2Var(\widehat{\mu}_2^2)\right) + O(n^{-4}). \tag{33}$$

□

The next theorem discusses $MSE(V^\times)$ in general.

Theorem 3 *Let $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F$ and $E(X^8) < \infty$. If $K_{G(.|\mathcal{X})}$ is independent of the observations, then by using the bootstrap estimation given in (i) and (ii) in Section 2*

$$MSE(V^*) = \left(\frac{B-1}{B}\right)^2 V(\widehat{V}(S_X^2)) + (E(V^*) - V(S_X^2))^2, \tag{34}$$

$$\begin{aligned} MSE(V^\#) = & \left(\frac{B-1}{B}\right)^2 \left(\frac{n-1}{n^3}\right)^2 ((n-1)K_{G(.|\mathcal{X})} - (n-3))^2 V(S_X^4) \\ & + (E(V^*) - V(S_X^2))^2, \end{aligned} \tag{35}$$

where $E(V^*)$ and $E(V^\#)$ are given in Theorem 2 and $V(\widehat{V}(S_X^2))$ can be found by Lemma 4.

Proof: The proof can be obtained directly by using the definition of MSE, Lemma 3 and Lemma 4. □

This theorem can be used to find $MSE(V^\times)$ for the nonparametric and parametric bootstrap. The next corollary is an application of this theorem for the normal distribution.

Table 1: Data used to study the simulation of variance

Variable	K_{F_n}	Data
x	2.59	48 36 20 29 42 42 20 42 22 41 45 14 6
		0 33 28 34 4 32 24 47 41 24 26 30 41
y	3.41	48 36 20 29 42 42 20 42 22 41 45 14 30
		0 33 28 34 24 32 24 47 41 24 26 30 41

Corollary 2 *Let* $\mathcal{X} = (X_1, \cdots, X_n) \overset{iid}{\sim} F = N(\mu, \sigma^2)$ *and if* $G(.|\mathcal{X}) = N(\bar{X}, S_X^2)$*, then the following relation holds asymptotically,*

$$MSE(V^{\#}) < MSE(V^*). \tag{36}$$

PROOF: This can be shown by using Corollary 1 and Theorem 3. □

Corollary 2 holds when F and G are normally distributed. However, if they have different distributions, it is rather difficult to conclude the same results and these should be studied by Theorem 3.

4 Simulations

This section includes the simulations of the bootstrap methods to clarify the results which are obtained theoretically in Section 3.

Example 3. Let the distribution function of the parametric bootstrap be $N(0, 1)$ where $K_{G(.|\mathcal{X})} = 3$. The chosen data set is the spatial perception of 26 neurologically impaired children. These data are used by Efron and Tibshirani (1993) to study the variance. The data set y is the same as x with a few changes to increase the kurtosis from 2.59 to 3.41 which helps to study the effect of kurtosis on the bootstrap estimation of variance. The data sets are given in Table 1.

Columns 2 and 4 in Table 2 include $S^{2\times}$ and $V^{\times}$ which are carried out with $B = 500$. Monte Carlo simulation (1000 replications) is used to study the comparison of nonparametric and parametric bootstrap estimation. The results of these Monte Carlo simulations are given in Column 3 and 5, as the ratios $S^{2*} < S^{2\#}$ and $V^* < V^{\#}$.

They agree with the theoretical analysis discussed in Section 3. The estimations of S^{2*} and $S^{2\#}$ are closed, but in contrast V^* and $V^{\#}$ are quite different. For the first data set with $K_{F_n} = 2.59$, with the assumption of normality of data, the nonparametric estimation of the variance bootstrap

Table 2: Simulation of $S^{2\times}$ and $V^{2\times}$

Variable	$S^{2*}, S^{2\#}$	Ratio†	$V^*, V^\#$	Ratio‡
x	165.00,165.00	0.517	1755.70,2178.04	0.990
y	118.31,118.37	0.492	1338.85,1124.60	0.042

† The ratio of $S^{2*} < S^{2\#}$ in the 1000 simulations,

‡ The ratio of $V^* < V^\#$ in the 1000 simulations.

method is more concentrated than the parametric estimation, which it can be seen from the value V^* and $V^\#$ in column 4 for the first data set. However the estimators of S^{2*} and $S^{2\#}$ are close not only for the first but also for the second data set. For the first data set, 99% of the simulations have $V^* < V^\#$. For the second data set, the parametric bootstrap is more concentrated than the nonparametric bootstrap, as only 42 out of 1000 have $V^* < V^\#$. This result agrees with the theoretical discussion of Theorem 1, because K_{F_n} of the second data is larger than 3.

Example 4. Here confidence intervals of variance are briefly discussed. There are different methods to find the confidence interval of variance which are discussed by Efron and Tibshirani (1993) and Davison and Hinkley (1997).

Method I The common method uses the CLT:

$$S^2 \pm t_{\alpha/2,n-1} S^2 \sqrt{\frac{(\widehat{K}-1)}{n}}.$$

Method II Letting $X_i \sim N(\mu, \sigma^2)$, it is easy to find the CI for variance:

$$\Big(\frac{nS^2}{\chi^2_{\alpha/2,n-1}}, \frac{nS^2}{\chi^2_{1-\alpha/2,n-1}}\Big).$$

Method III This method is referred to as the standard method:

$$S^2 \pm t_{\alpha/2} \widehat{se}^\times_B,$$

where $\widehat{se}^\times_B$ is the bootstrap estimate of standard error.

Method IV The CI of variance based on the bootstrap methods can be found by using the following formula:

$$S^2 \pm t^\times_{\alpha/2} S^2 \sqrt{\frac{(K-1)}{n}},$$

where $t^{\times}_{\alpha/2}$ is $\alpha/2$ percentile of $t^{\times} = \dfrac{S^{2\times} - S^2}{\sqrt{V(S^2)^{\times}}}$, where $S^{2\times}$ and $V(S^2)^{\times}$ are estimated by the bootstrap method.

Method V Using method II, the CI is asymptotically as below:

$$\left(\frac{nS^2}{\chi^{2\times}_{\alpha/2}}, \frac{nS^2}{\chi^{2\times}_{1-\alpha/2}}\right),$$

where $\chi^{2\times}_{\alpha/2}$ is the percentile of $\chi^{2\times} = \dfrac{nS^{2\times}}{S^2}$.

Method VI This method is called the percentile CI:

$$[\widehat{\theta}_{\%low}, \widehat{\theta}_{\%up}] = [\widehat{G}^{-1}(\alpha/2), \widehat{G}^{-1}(1-\alpha/2)],$$

where $\widehat{G}^{-1}(\alpha/2) = S^{2\times}(\alpha/2)$, the percentile of the bootstrap resampling of variance.

Method VII This method, referred to as bias-corrected and accelerated, BC_{α}, by Efron and Tibshirani (1993), has a substantial improvement over the percentile method in both theory and practice. It is based on the percentile of bootstrap distribution by adjusting the acceleration and bias-correction.

Tables 3 and 4 include the bootstrap confidence interval at the level of 95% for x and y with $B = 500$ which are discussed in the previous example. The parametric bootstrap is done by the normal distribution. The first two lines of both tables are the standard method for the construction of CI of variance, which is based on the t and χ^2. It is obvious that Method I has smaller length than Method II because the former is based on the symmetrical distribution but in reality the distribution of variance is asymmetrical. Method II is known as the exact method, as a criterion which can be used to study the different methods.

It is obvious that the length of parametric bootstrap CI for the variance of x is wider than the nonparametric bootstrap. In contrast the length of parametric bootstrap CI for the variance of y is shorter than that of the nonparametric bootstrap. Because Method III uses the square root of $V(S^2)$ which depends on the kurtosis, this method is directly affected by kurtosis. Method IV uses bootstrap resamples in t. Although the methods V and VI do not use $V(S^2)^{\times}$ directly, they are based on the 5th and 95th percentiles and of course the spread is directly affected by kurtosis. Method VII is also

Table 3: Confidence interval at 95% for x and y

	x			y		
Method	Low	Up	Length	Low	Up	Length
I	99.018	244.049	145.031	59.050	187.094	128.043
II	118.448	305.233	186.784	84.984	218.999	134.014
III non	100.064	243.003	142.938	61.886	184.258	122.372
III par	91.483	251.584	160.101	66.084	180.060	113.976
IV non	110.249	283.828	173.578	75.743	295.122	219.379
IV par	115.847	309.760	193.912	77.494	236.850	159.356
V non	124.379	306.281	181.902	83.498	225.435	141.936
V par	120.598	311.475	190.876	85.460	219.482	134.022
VI non	99.927	233.364	133.437	65.223	183.094	117.870
VI par	96.051	248.405	152.353	69.262	175.156	103.660
VII non	119.520	258.307	138.786	79.792	227.236	147.443
VII par	113.565	289.907	176.342	84.531	217.723	133.192

Table 4: Simulation of convergence of V^* and $V^\#$ to $V(S_X^2)$

	$K_{F_n} < 3$		$K_{F_n} > 3$	
Size	Ratio1	Ratio2	Ratio1	Ratio2
$n = 10$	0.718	0.282	0.403	0.597
$n = 30$	0.695	0.305	0.418	0.518
$n = 50$	0.676	0.324	0.562	0.438
$n = 100$	0.676	0.324	0.581	0.418

Ratio1: Number of times $|V(S_X^2) - V^\#| < |V(S_X^2) - V^*|$, Ratio2=1-Ratio1

affected by it, see complete discussion in section 2.2.

Example 5. This example explains the convergence of V^* and $V^\#$ to $V(S_X^2)$ by simulation. Let F be the distribution of $N(0,1)$ and have the parametric bootstrap done by the normal distribution with B=500. Table 4 shows how many times the bootstrap estimation of standard error is close to $V(S_X^2)$ for the results based on the 1000 simulations.

Table 4 shows that when $K_{F_n} < 3$, the parametric bootstrap method is appropriate although the replications of nonparametric bootstrap concentrate more but when $K_{F_n} > 3$ and n is small then the nonparametric bootstrap has better performance. This results admit Corollary 1 and its discussions.

5 Conclusions

This paper discusses bootstrap estimation of the variance, in the nonparametric and the parametric setting and studies their behavior. It shows that the parametric and nonparametric bootstrap estimation of variance are equal (7), but that the bootstrap standard error depends on the sample kurtosis (7). If the distribution of the sample is normal and the parametric bootstrap is based on the normal distribution then this parametric bootstrap method with normal distribution can be expected to be better than the nonparametric bootstrap, i.e. closer to the sample distribution. If $K_{F_n} > 3$, then for small sample sizes, the nonparametric bootstrap method seems more appropriate.

Moreover, Theorem 2 gives the expectation of V^* and $V^{\#}$. In the case of the nonparametric method, this depends on K but for the parametric method, it depends on K and $K_{G(.|\mathcal{X})}$. When $K_{G(.|\mathcal{X})}$ depends on the observations, the given general form of the parametric bootstrap does not hold.

Figure 2 explains the expected result. It clarifies that when K is between 1.4 and 2, the result of nonparametric bootstrap is more appropriate, regardless of whether $K_{G(.|\mathcal{X})}$ and F have the same distribution.

This paper emphasizes that special care should be taken when making claims about the accuracy of the parametric bootstrap approach in applications, Figure 3, which is based on Theorem 2, clarifies how much the result is affected by the wrong choice if the distribution of population is not normal.

Two kind of expectations are discussed throughout, conditional and unconditional. The conditional expectation clarifies the result of the bootstrapping, whereas the unconditional expectation is the combination of the bootstrapping and a frequentist approach.

References

[1] Athreya, K.B. and Lahiri. S.N. (2006). *Measure Theory and Probability Theory.* Springer, New York.

[2] Chiodi, M. (1995). Generation of pseudo random variates from a normal distribution of order P. *Pubblicato su Statistica Applicata.* **7**(4), 401-416.

[3] Cramér, H. (1945). *Mathematical Methods of Statistics.* Almqvist & Wiksells, Uppsala.

[4] Darlington, R.B. (1970). Is kurtosis really peakedness? *The American Statistician.* **24**, 19-22.

[5] Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application.* Cambridge University Press. Cambridge.

[6] Efron, B. and Tibshirani, R. (1993). *Introduction to the Bootstrap.* Chapman & Hall, New York.

[7] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.
[8] Hintze, J.L. and Nelson, R.D. (1998). Violin plots: a box plot-density trace synergism. *The American Statistician*, **2**(5), 181-4.
[9] Janssen, A. and Pauls,T. (2003). How do bootstrap and permution tests work? *Annals of Statistics*, **31**, 768-806.
[10] Joanes, D.N. and Gill, C.A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician*, **47**(1), 183-189.
[11] Lee, S.S. (1994). Optimal choice between parametric and non-parametric bootstrap estimates. *Math. Proc. Comb. Phil. Soc*, **115**, 335.
[12] Mackinnon, J.G. (2002). Bootstrap inference in econometrics. *Canadian Journal of Economics*, **35**, 615-645.
[13] Ostaszewski, K. and Rempala, G.A. (2000). Parametric and nonparametric bootstrap in acturial practice. *www.actuarialfoundation.org/research_edu/parametic.pdf.*
[14] Shao, J. and Tu, D. (1996). *The Jackknife and Bootstrap*. Springer, New York.