A Coverage Probability of Bootstrap-t Confidence Interval for the Variance

Vesna Ć. Rajić

Abstract. We examine one-sided confidence intervals for the population variance, based on the ordinary $t$-statistics. We derive an unconditional coverage probability of the bootstrap-$t$ interval for unknown variance. For that purpose, we find an Edgeworth expansion of the distribution of $t$-statistic to an order $n^{-2}$. We can see that a number of simulation, $B$, has the influence on coverage probability of the confidence interval for the variance. If $B$ equals sample size then coverage probability and its limit (when $B \to \infty$) disagree at the level $O(n^{-2})$. If we want that nominal coverage probability of the interval would be equal to $\alpha$, then coverage probability and its limit agree to order $n^{-2}$ if $B$ is of larger order than the square root of the sample size. We present a modeling application in insurance property, where the purpose of analysis is to measure variability of a data set.

1. Introduction

Hall [10] in his paper gave some conclusions about the effect of the number of bootstrap simulations on the bootstrap-$t$ confidence intervals. One of that conclusions concerns coverage probability in case of application to smooth statistics, such as the Studentized mean of a sample drawn from a continuous distribution. He also gave an explicit formula for the second-order term in an expansion of coverage probability for the case of Studentized mean. In this paper we shall examine one-sided confidence intervals for the population variance based on the ordinary $t$ statistics. We shall derive an unconditional coverage probability of bootstrap-$t$ confidence interval for unknown variance based on a sample from a continuous distribution. From that formula it will be possible to make some points about the number of bootstrap simulations required to construct a bootstrap-$t$ confidence interval for population variance. Bootstrap-$t$ confidence intervals for the variance are considered in [6] and [7] and these intervals can be used for various modeling applications.

This paper will be organized as follows. In Section 2 we shall derive an Edgeworth expansion for one sample $t$-statistic (that will be used for estimation variance) to order $n^{-2}$. Edgeworth expansion of Students statistic was investigated by several authors (see [3], [5], [13], [15], [17], [18]). In Section 3 we shall briefly mention bootstrap-$t$ intervals. Properties of those intervals were investigated in a series of papers ([11], [2], [14], [16]). In this section we shall give an explicit formula for the second-order term in an expansion of coverage probability for the case of Studentized variance. In Sections 4 and 5 we shall conduct
a simulation study to assess the coverage accuracy of presented confidence intervals. In Section 6 we shall give concluding remarks.

2. Edgeworth Expansion for the Studentized Variance

Let \( X_1, ..., X_n \) be i.i.d. from normal distribution with mean \( \mu \) and variance \( \sigma^2 \). It is known that statistic
\[
\frac{(n-1)S^2}{\sigma^2}
\]
has \( \chi^2_{n-1} \) distribution, where \( S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \) is a sample variance. From the fact that for enough large \( n \) \( \chi^2_{n-1} \) distribution can be approximated by normal, follows that the distribution of the variable:
\[
Z = \frac{(n-1)S^2 - (n-1)}{\sqrt{2(n-1)}} = \frac{S^2 - \sigma^2}{\sqrt{\text{var} (S^2)}}
\]
converges to standardize normal distribution as \( n \) increases to infinity (for details see [7]). Let us consider statistic
\[
T = \frac{S^2 - \sigma^2}{\sqrt{\text{var} (S^2)}},
\]
where \( \sqrt{\text{var} (S^2)} \) is a consistent estimator of the variance of \( S^2 \). We shall derive an Edgeworth expansion for above mentioned \( t \)-statistic. Before that, let define random variables \( X_i' = \frac{(X_i - \overline{X}) - \frac{n-1}{n} \sigma^2}{\sqrt{\text{var} (X_i)}} \) for \( i = 1, 2, ..., n \), where \( V_1 = E \left( \left( X_i' \right)^2 - \frac{n-1}{n} \sigma^2 \right)^2 \).

Proposition 2.1. If Cramer’s condition holds (see [14]) and if \( EX_i^{10} < \infty \), the distribution of \( t \) statistic given in equation (2.1) has the following expansion
\[
P(T \leq x) = \Phi(x) + \frac{1}{\sqrt{n}} q_1(x) \phi(x) + \frac{1}{n} q_2(x) \phi(x) + \frac{1}{n \sqrt{n}} q_3(x) \phi(x) + O \left( \frac{1}{n^2} \right),
\]
where functions \( q_1(x), q_2(x), q_3(x) \) have the form:
\[
q_1(x) = \frac{M_3^2}{6} \left( 2x^2 + 1 \right),
q_2(x) = \left( \frac{1}{3} + \frac{3}{8} M_4 - \frac{1}{8} M_6^2 \right) x
\]
\[
q_3(x) = -\frac{1}{4} M_3^2 - \frac{1}{6} \left( \frac{3}{2} M_3^2 - \frac{1}{4} M_5^2 + 3 M_3' M_4' + \frac{3}{4} M_5' \right) (x^2 - 1).
\]
\( \phi(\cdot) \) and \( \Phi(\cdot) \) are the probability density function and cumulative distribution function of standard normal variable and \( M_k^* = E \left( \frac{1}{n} \sum_{i=1}^{n} X_i^k \right) \), \( (k = 3, 4, 5) \). For the proof see Appendix A.

3. Coverage Probability of One-Sided Confidence Interval for the Variance

Let \( S^2 \) be an estimator of a parameter \( \sigma^2 \), based on a random sample \( (X_1, ..., X_n) \). We shall consider a bootstrap-\( t \) confidence interval for \( \sigma^2 \) based on the statistic (2.1). After generating \( B \) bootstrap samples, in each bootstrap sample we compute the value of statistic:
\[
T^* = \frac{S^2 - \hat{S}^2}{\sqrt{\text{var} (S^2)}},
\]
where \( S^2 \) is a bootstrap replication of \( S^2 \).
Let denote by $t_{α}$ the point which is the bootstrap approximation to the point $x_{α}$, such that $P(T ≤ x_{α}) = α$. Let $T_{1}^{*}, T_{2}^{*}, ..., T_{B}^{*}$ be independent copies of statistic $T^{*}$ in $B$ bootstrap samples, arranged in ascending order. If we select $T_{(v+1)}$ as an approximation to $t_{α}$ ($0 ≤ v ≤ B-1$) then the exact, unconditional coverage probability of $I_{boot} = [s^{2} - T_{(v+1)}^{∗}, √{var(s^{2})}, +∞)$, is

$$
α (v, B) = \frac{v + 1}{B + 1} - \frac{1}{n} \int_{0}^{1} R_{n} (c) \, dc,
$$

(2.3)

where $R_{n}$ is such function that bounded uniformly in $n ≥ 1$ and $0 < α < 1$. The point $τ = τ(ν, ν)$ is the solution of equation $G(τ) = ν$, for $0 < ν < 1$, where $G(ν) = ∑_{j=0}^{ν} (\binom{ν}{j}) (1 - u)^{B-j}$. Asymptotic formula for $R_{n} (α)$ follows from (see [10]):

$$
P (p ≤ α) = α + \frac{1}{n} R_{n} (α),
$$

where $p = P (T^{∗} < T | X)$ and Hall in [11] investigated that formula in case of Studentized mean. We shall concentrate on a case of Studentized variance to get coverage probability of the interval boot $I_{boot}$. For that purpose we find an asymptotic formula for $R_{n} (α)$.

Whenever it exists, an Edgeworth expansion for the statistic (2.1) may be inverted to yield an expansion of (inverse) Cornish-Fisher type (see [10]):

$$
P \left( T ≤ x - \frac{1}{\sqrt{n}} q_{1} (x) - \frac{1}{n} q_{2} (x) - \frac{1}{n \sqrt{n}} q_{3} (x) \right) = \Phi (x) + O \left( n^{-2} \right).
$$

If in functions $q_{1} (x), q_{2} (x), q_{3} (x)$ corresponding moments $(M'_{i}, i = 3, 4, 5)$ replaced by its estimates based on the sample (note them by $m'_{i}, i = 3, 4, 5$), we get new functions which we shall denote by $q'_{1} (x), q'_{2} (x), q'_{3} (x)$. Asymptotic formula for $R_{n} (α)$ follows from expression:

$$
P \left( T ≤ x_{α} - \frac{1}{\sqrt{n}} q'_{1} (x_{α}) - \frac{1}{n} q'_{2} (x_{α}) - \frac{1}{n \sqrt{n}} q'_{3} (x_{α}) \right) = P (p ≤ α) + O \left( n^{-2} \right)
$$

$$
= α + \frac{1}{n} R_{n} (α) + O \left( n^{-2} \right),
$$

(2.4)

where $x_{α}$ is the solution of $Φ (x_{α}) = α$. To get this asymptotic formula we must find an Edgeworth expansion to order $n^{-2}$of the distribution $P (S(α) ≤ x)$, where statistic $S(α)$ is equal to:

$$
S(α) = T + \frac{1}{\sqrt{n}} q'_{1} (x_{α}) + \frac{1}{n} q'_{2} (x_{α}) + \frac{1}{n \sqrt{n}} q'_{3} (x_{α}).
$$

(2.5)

Functions $Q_{1} (x), Q_{2} (x)$ and $Q_{3} (x)$ have the form:

$$
Q_{1} (x) = \frac{M'_{3}}{3} (x^{2} - z_{α}^{2}) \phi (x).
$$

$$
Q_{2} (x) = \left( \frac{1}{8} z_{α} \left( - \frac{3}{8} M'_{4} z_{α} + \frac{1}{8} M'_{3}^{2} z_{α} - \frac{1}{2} x \left( \frac{1}{4} M'_{3}^{2} + M'_{4} \left( \frac{2}{3} z_{α}^{2} - \frac{5}{12} \right) - \frac{1}{4} \right) \right) \phi (x) \right),
$$
If we set $x = z_a$ in (2.5) we get:

$$P(S(a) \leq z_a) = \alpha - \frac{1}{6n^2} M_4 z_a (2z_a^2 + 1) \phi(z_a) + \frac{1}{n \sqrt{n}} \left( -\frac{3}{2} M_5 z_a^2 + \frac{1}{6} M_3 M_4' \left( z_a^4 + z_a^2 - \frac{1}{2} \right) \right) \phi(z_a) + O \left( \frac{1}{n^2} \right).$$

(2.6)

From (2.6) follows asymptotic formula for $R_n(a)$:

$$R_n(a) = \varphi_1(z_a) \phi(z_a) + \frac{1}{n \sqrt{n}} \varphi_2(z_a) \phi(z_a) + O \left( \frac{1}{n^2} \right),$$

where $\varphi_1(z_a) = -\frac{1}{6} M_4^2 z_a (2z_a^2 + 1)$, $\varphi_2(z_a) = -\frac{1}{6} M_5^2 z_a^2 + \frac{1}{6} M_3^2 M_4' \left( z_a^4 + z_a^2 - \frac{1}{2} \right)$. Now we have

$$P(p \leq a) = \alpha + \frac{1}{n} \Psi_1(a) + \frac{1}{n \sqrt{n}} \Psi_2(a) + O \left( \frac{1}{n^2} \right),$$

where $\Psi_i(a) = \varphi_i(z_a) \phi(z_a)$, $i = 1, 2$. From (2.3) follows:

$$\alpha(v, B) = \frac{v + 1}{B + 1} + \frac{1}{n} \int_0^1 \Psi_1(\tau) d\tau + \frac{1}{n^2} \int_0^1 \Psi_2(\tau) d\tau + O \left( \frac{1}{n^2} \right),$$

for $0 \leq v \leq B - 1$ and $B \geq 1$. Value $\tau$ can be find from equation $G(\tau) = v$ and is of order $o \left( \frac{1}{B^{-1}} \right)$. If we apply the mean-value theorem for integrals, we get a coverage probability of one-sided bootstrap–$t$ interval for variance, which is equals to:

$$\alpha(v, B) = \alpha' + \frac{1}{n} \Psi_1(\alpha') + \frac{1}{n \sqrt{n}} \Psi_2(\alpha') + O \left( \frac{1}{n^3} \right),$$

where $\alpha' = \frac{v + 1}{B + 1}$ is the nominal coverage probability of the confidence interval $l_{\text{boot}}$.

4. A Simulation Study

In this Section we investigate the coverage accuracy of one-sided bootstrap–$t$ confidence intervals for population variance. The distributions that we consider are Normal and one-parameter Weibull. The whole approach can be applied on any other distribution. For each sample size (10, 20, 50, 80, 100, 200, 500), we decided to generate a fixed number of bootstrap samples (1000, 5000). Results of coverage of 95% confidence intervals are presented in Table 1.
Table 1: Coverage probability of 95% one-sided confidence intervals for the variance of different distributions

<table>
<thead>
<tr>
<th>n</th>
<th>B</th>
<th>N(0,1)</th>
<th>Weibull(0.5)</th>
<th>Weibull(1)</th>
<th>Weibull(2)</th>
<th>Weibull(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1000</td>
<td>0.929</td>
<td>0.994</td>
<td>0.975</td>
<td>0.940</td>
<td>0.931</td>
</tr>
<tr>
<td>20</td>
<td>1000</td>
<td>0.937</td>
<td>0.987</td>
<td>0.968</td>
<td>0.941</td>
<td>0.932</td>
</tr>
<tr>
<td>50</td>
<td>1000</td>
<td>0.945</td>
<td>0.985</td>
<td>0.959</td>
<td>0.956</td>
<td>0.933</td>
</tr>
<tr>
<td>80</td>
<td>1000</td>
<td>0.948</td>
<td>0.982</td>
<td>0.958</td>
<td>0.944</td>
<td>0.942</td>
</tr>
<tr>
<td>100</td>
<td>5000</td>
<td>0.943</td>
<td>0.976</td>
<td>0.956</td>
<td>0.947</td>
<td>0.938</td>
</tr>
<tr>
<td>200</td>
<td>1000</td>
<td>0.948</td>
<td>0.979</td>
<td>0.957</td>
<td>0.945</td>
<td>0.942</td>
</tr>
<tr>
<td>500</td>
<td>1000</td>
<td>0.951</td>
<td>0.969</td>
<td>0.944</td>
<td>0.944</td>
<td>0.942</td>
</tr>
<tr>
<td>100</td>
<td>5000</td>
<td>0.950</td>
<td>0.970</td>
<td>0.954</td>
<td>0.947</td>
<td>0.946</td>
</tr>
<tr>
<td>500</td>
<td>1000</td>
<td>0.949</td>
<td>0.959</td>
<td>0.950</td>
<td>0.951</td>
<td>0.943</td>
</tr>
<tr>
<td>100</td>
<td>5000</td>
<td>0.949</td>
<td>0.959</td>
<td>0.948</td>
<td>0.949</td>
<td>0.953</td>
</tr>
</tbody>
</table>

From Table 1 it can be seen that the bootstrap-t interval for the population variance gives good coverage when sample comes from Normal distribution. If we have sample from Weibull distribution, this method gives good coverage when parameter of Weibull distribution is greater than 1, even for small sample size. Method is implemented using programming language Fortran.

5. A Real Example

Here, we analyze risk management of losses in tariff glass breakage insurance. Actually, we estimate the variance by using real data set about incurred losses. We got data from one insurance company in Serbia for 2014. The descriptive statistics for the data set are given in the Table 2. Figure 1 shows the histogram of losses.

<table>
<thead>
<tr>
<th>Data set</th>
<th>N</th>
<th>Mean</th>
<th>Std. deviation</th>
<th>Skewness coef.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tariff glass breakage insurance</td>
<td>552</td>
<td>4438.92</td>
<td>6555.923</td>
<td>4.05</td>
</tr>
</tbody>
</table>

Figure 1: Histogram of data set
By measuring loss variability we will be able to derive precious conclusions and to determine adequate premium principle. Frequently used premium principle in property is Standard Deviation Premium Principle (see [4]), which includes a risk load that is proportional to the standard deviation of the risk (see [4] and [7]). Because of that, it is very important to estimate variance of data. For that purpose we investigate the coverage accuracy of one-sided bootstrap-t confidence intervals for population variance. From the above-mentioned data set from we generated samples of size 10, 20, 50, and 100 and from each sample we generated 1000 bootstrap samples. Results are presented in Table 3.

Table 3: Coverage probability of 95% confidence intervals for the variance

<table>
<thead>
<tr>
<th>Sample size</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage probability</td>
<td>0.678</td>
<td>0.835</td>
<td>0.902</td>
<td>0.975</td>
</tr>
</tbody>
</table>

From Table 3 it is obvious (because of skewed distribution) that bootstrap-t intervals have the good coverage accuracy for large sample size. Therefore, for measuring variability in insurance property we recommend using of these intervals.

6. Concluding Remarks

In this paper we have shown how can be constructed confidence intervals for unknown population variance, using ordinary t-statistic. It is possible for large sample size, because in that case $\chi^2$ distribution can be approximated by normal distribution. Based on the $t$ statistic is possible to construct bootstrap-t interval for the variance which is second-order accurate (see [16]) and has a lot of nice characteristics. Bootstrap-t procedure, by now, has been almost applied to location statistics, like the sample mean, median, trimmed mean or a sample percentile. We suggested its applying to sample variance. For that purpose, we found an Edgeworth expansion of the distribution of mentioned t statistic (to order $n^{-2}$) and used that expansion to find an explicit formula for coverage probability of one-sided bootstrap-t interval. We saw that a number of simulation, $B$, had the influence on coverage probability of one-sided confidence interval for population variance. If $B$ equals sample size then $a(v, B)$ and its limit $P(p \leq a)$ (when $B \to \infty$) disagree at the level $O(n^{-2})$. If we want that nominal coverage probability of $I_{boot}$ would be equal to $\alpha$, then $a(v, B)$ and its limit $P(p \leq a)$ agree to order $n^{-2}$ if $B$ is of larger order than the square root of the sample size. Those conclusions, about the number of bootstrap simulation, are the same with conclusions that made Hall[11] in case of one-sided confidence interval for population mean.

Appendix A.

Here we prove Proposition 1. We kept some notations from [7] and [18]. The $t$-statistic is given by

$$T = \frac{\sqrt{\frac{\sum_{i=1}^{n} X'_i}{n-1}} \cdot \frac{\sum_{i=1}^{n} Y'_i}{\sqrt{n-1} \cdot \sqrt{\sum_{i=1}^{n} X'^2_i}}}{\sqrt{\frac{\sum_{i=1}^{n} X'_i}{n-1} \cdot \frac{\sum_{i=1}^{n} Y'_i}{\sqrt{\sum_{i=1}^{n} X'^2_i}}}} = \sqrt{n-1} \cdot g(Y),$$

where $g(Y) = \frac{\sqrt{n}}{\sqrt{Y_2 - Y_1}}$. Let $Y \equiv (Y_1, Y_2)$ and $EY \equiv U \equiv (U_1, U_2) = (0, 1)$. We define statistic $W_n$ as follows:

$$W_n = \sqrt{n-1} \left( \frac{\partial g}{\partial Y_1} (U) \cdot (Y_1 - U_1) + \frac{\partial g}{\partial Y_2} (U) \cdot (Y_2 - U_2) + \frac{1}{2} \left[ \frac{\partial^2 g}{\partial Y_1^2} (U) \cdot (Y_1 - U_1)^2 + 2 \frac{\partial^2 g}{\partial Y_1 \partial Y_2} (U) \cdot (Y_1 - U_1)(Y_2 - U_2) + \frac{\partial^2 g}{\partial Y_2^2} (U) \cdot (Y_2 - U_2)^2 \right] \right).$$

V. Č. Rajić / Filomat 31:10 (2017), 2967–2974
2972
After some calculations it becomes: \( W_n = \sqrt{n-1} \left( \frac{1}{2} Y_1 - \frac{1}{2} Y_1 Y_2 \right) \).

If we used \( \sqrt{n-1} = \sqrt{n} - \frac{1}{2\sqrt{n}} + o \left( \frac{1}{n} \right) \), the first three moments of statistic \( W_n \) are:

\[
EW_n = -\frac{1}{2} \frac{1}{\sqrt{n}} M'_3 + \frac{1}{4n} \frac{1}{\sqrt{n}} M'_5 + O \left( \frac{1}{n^2} \right),
\]

\[
EW^2_n = 1 - \frac{3}{4n} M'_4 + \frac{1}{2n} M'_4^2 - \frac{11}{4n} + O \left( \frac{1}{n^2} \right),
\]

\[
EW^3_n = -\frac{7}{2} \frac{1}{\sqrt{n}} M'_3 + \frac{1}{n} \left( \frac{21}{8} M'_4 + \frac{33}{8} M'_4 M'_4 + \frac{3}{4} M'_5 - \frac{3}{4} M'_5^3 \right) + O \left( \frac{1}{n^2} \right).
\]

Let \( \psi_n (t) \) be the characteristic function of \( W_n \). Then

\[
\psi_n (t) = \exp \left\{ K_{1n} (it) + \frac{1}{2} K_{2n} (it)^2 + \frac{1}{6} K_{3n} (it)^3 + \cdots \right\},
\]

where the first three cumulants of statistic \( W_n \) are:

\[
K_{1n} = -\frac{1}{2} \frac{1}{\sqrt{n}} M'_3 + \frac{1}{8n} \frac{1}{\sqrt{n}} M'_5 + O \left( \frac{1}{n^2} \right),
\]

\[
K_{2n} = EW^2_n - (EW_n)^2 = 1 + \frac{1}{8n} \left( -1 - 3M'_4 + M'_2^2 \right) + O \left( \frac{1}{n^2} \right),
\]

\[
K_{3n} = E( W_n - EW_n)^3 = -\frac{2}{8n} M'_3 + \frac{1}{4n} \left( \frac{3}{2} M'_4 - \frac{1}{4} M'_5^3 + 3M'_4 M'_4 + \frac{3}{4} M'_5 \right) + O \left( \frac{1}{n^2} \right).
\]

The characteristic function \( \psi_n (t) \) has the form:

\[
\psi_n (t) = \exp \left\{ \frac{-t^2}{2} \left[ 1 + \frac{1}{\sqrt{n}} \left( \frac{-1}{2} M'_3 (it) - \frac{2}{6} M'_4 (it)^3 \right) + \frac{1}{n} \left( \frac{-1}{8} M'_4 + \frac{1}{2} M'_5 \right) (it)^2 \right. \right.
\]

\[
\left. + \frac{1}{\sqrt{n}} \left( \frac{1}{4} M'_3 (it) + \frac{1}{6} \frac{3}{2} M'_4 - \frac{1}{4} M'_5 + 3M'_4 M'_4 + \frac{3}{4} M'_5 \right) + O \left( \frac{1}{n^2} \right) \right]\right\} + O \left( \frac{1}{n^2} \right).
\]

Since we have \( \psi_n (t) = \int_{-\infty}^{+\infty} e^{itx} dP (W_n \leq x) \) and \( e^{-t} = \int_{-\infty}^{+\infty} e^{itx} d\Phi (x) \), above expression becomes:

\[
P (W_n \leq x) = \Phi (x) + \frac{1}{\sqrt{n}} R_1 (x) + \frac{1}{n} R_2 (x) + \frac{1}{n \sqrt{n}} R_3 (x) + O \left( \frac{1}{n^2} \right).
\]

Function \( R_i (x), i = 1, 2, 3 \) can be calculated from integrals: \( \int_{-\infty}^{+\infty} e^{itx} dR_i (x) = r_i (it) e^{-\frac{t^2}{2}} \), where

\[
r_1 (it) = -\frac{1}{2} M'_3 (it) - \frac{2}{6} M'_4 (it)^3, \quad r_2 (it) = \left( \frac{-1}{2} M'_4 + \frac{1}{8} M'_5 \right) (it)^2,
\]

\[
r_3 (it) = \frac{1}{4} M'_3 (it) + \frac{1}{6} \left( \frac{3}{2} M'_4 - \frac{1}{4} M'_5^3 + 3M'_4 M'_4 + \frac{3}{4} M'_5 \right) (it)^3.
\]

Since, \( T = W_n + O \left( \frac{1}{n^1} \right) \), we have: \( P (T \leq x) = \Phi (x) + \frac{1}{\sqrt{n}} R_1 (x) + \frac{1}{n} R_2 (x) + \frac{1}{n^{1/2}} R_3 (x) + O \left( \frac{1}{n} \right) \), where

\[
R_1 (x) = \frac{M'_3}{6} \left( 2x^2 + 1 \right) \phi (x), \quad R_2 (x) = \left( \frac{1}{8} M'_4 - \frac{1}{8} M'_5 \right) x \phi (x),
\]

\[
R_3 (x) = \left( -\frac{1}{6} M'_3 + \frac{1}{3} \frac{3}{2} M'_4 - \frac{1}{4} M'_5^3 + 3M'_4 M'_4 + \frac{3}{4} M'_5 \right) (x^2 - 1) \phi (x).
\]
Appendix B.

The first three moments of the statistic $S(\alpha)$ given in equation (2.4) are:

$$E(S(\alpha)) = \frac{1}{3} \frac{1}{n} M_3'(z_a^2 - 1) + \frac{1}{n} \left( \frac{1}{8} z_a + \frac{3}{8} M_3' z_a - \frac{1}{8} M_3'' z_a \right)$$

$$+ \frac{1}{n \sqrt{n}} \left( -\frac{1}{4} M_5'(z_a^2 - 1) + \frac{1}{24} M_5''(z_a^2 - 1) - \frac{1}{2} M_5'M_4'(z_a^2 - 1) - \frac{1}{8} M_5'(z_a^2 - 1) \right) + O(n^{-2})$$

$$E(S(\alpha))^2 = 1 + \frac{1}{n} \left( M_3'^2 \left( \frac{1}{9} z_a - \frac{2}{9} z_a^2 + \frac{13}{36} \right) + M_4' \left( \frac{2}{3} z_a^2 - \frac{5}{12} \right) - \frac{1}{4} \right)$$

$$+ \frac{1}{n \sqrt{n}} \left( \frac{3}{8} z_a M_5' + \frac{1}{12} M_3'(z_a^2 - z_a) - \frac{1}{12} M_3^2(z_a - z_a) + M_3'M_4' \left( \frac{1}{4} z_a^2 - \frac{3}{4} z_a \right) \right) + O(n^{-2})$$

$$E(S(\alpha))^3 = \frac{1}{n \sqrt{n}} M_3'(z_a - 3) + \frac{1}{n} \left( \frac{3}{8} z_a + \frac{9}{8} z_a M_4' - \frac{3}{8} z_a M_3^2 \right)$$

$$+ \frac{1}{n \sqrt{n}} \left( M_3'^2 \left( \frac{5}{2} - z_a^2 \right) + M_4' \left( \frac{9}{8} - \frac{3 z_a^2}{8} \right) \right)$$

$$+ M_5' \left( \frac{1}{27} z_a^6 - \frac{1}{9} z_a^4 + \frac{35}{72} z_a^2 - \frac{143}{216} \right) + M_5'M_4' \left( \frac{2}{3} z_a^4 - \frac{43 \alpha}{12} z_a^2 + \frac{53}{12} \right) + O(n^{-2})$$

From this moments is easy to find the cumulants of the variable $S(\alpha)$ and then use that cumulants to obtain an Edgeworth expansion of $P(S(\alpha) \leq x)$. Procedure is the same like in appendix A.

References