Journal of Statistical Planning and Inference 141 (2011) 1667–1672

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

One-sided confidence intervals for population variances of skewed distributions

Vesna Cojbasic, Dragan Loncar *

Faculty of Economics, University of Belgrade, Kamenicka 6, 11000, Belgrade, Serbia

A R T I C L E   I N F O

Article history:
Received 15 October 2009
Received in revised form
6 November 2010
Accepted 6 November 2010
Available online 19 November 2010

Keywords:
Bootstrap
Edgeworth expansion
Variance
Revenue variability
Skewness

A B S T R A C T

In this article we study the coverage accuracy of one-sided bootstrap-*t* confidence intervals for the population variances combined with Hall's and Johnson's transformation. We compare the coverage accuracy of all suggested intervals and intervals based on the Chi-square statistic for variances of positively skewed distributions. In addition, we describe and discuss an application of the presented methods for measuring and analyzing revenue variability within the food retail industry. The results show that both Hall's transformation and Johnson's transformation approaches yield good coverage accuracy of the lower endpoint confidence intervals, better than method based on the Chi-square statistic. For the upper endpoint confidence intervals Hall's bootstrap-*t* method yields the best coverage accuracy when compared with other methods.

1. Introduction

When constructing a confidence interval for the variance, parametric method is based upon the χ^2-statistic (introduced by Pearson, 1900). $(1-\alpha)100\%$ one-sided confidence interval for an unknown parameter $\theta=\sigma^2$ is interval

$$\left(0, \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha}}\right), \tag{1}$$

where $\chi^2_{n-1;1-\alpha}$ is determined by $P(\chi^2_{n-1} > \chi^2_{n-1;1-\alpha}) = 1-\alpha, \chi^2_{n-1}$ is a random variable which has χ^2 distribution with $n-1$ degrees of freedom, and S^2 is a sample variance. If we call an interval (1) the upper endpoint interval, then the lower endpoint confidence interval for parameter σ^2 would be

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1;\alpha}}, +\infty\right). \tag{2}$$

These intervals are based on parametric assumptions that can be quite inaccurate in practice. If the assumptions are not fulfilled, then intervals based on the χ^2-statistic are highly sensitive (see Lehman, 1986). There does not seem to be much literature on constructing robust confidence intervals for the variance. Some interesting results on this topic were presented by Barham and Jeyaratnam (1999). They considered confidence intervals for the variance based on estimators which are much less sensitive to the presence of outliers.

* Corresponding author. Tel.: +381 113021075; mobile: +381 652405020.
E-mail address: loncar@ekof.bg.ac.rs (D. Loncar).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.11.007

In this paper we consider a new type of confidence intervals for the unknown population variance, named the bootstrap-t intervals. They are based on the ordinary t statistic, which can be modified in order to remove the skewness effects of a skewed distribution on the t statistic. So far, transformation of the t statistic was investigated by Johnson (1978), Abramovitch and Singh (1985), Hall (1992a), Chen (1995), Zhou and Gao (2000), Zhou and Dinh (2004) and other authors. Johnson (1978) proposed a transformation of the t statistic, but his transformation was not monotone and invertible. Hall (1992a) suggested a transformation aiming at eliminating skewness from the distribution of Studentized statistic. His transformation was monotone and had simple, explicit inversion formulae. Zhou and Gao (2000) studied the performance of both Hall's and Johnson's transformation methods and their bootstrapped versions. Zhou and Dinh (2004) suggested new transformed t statistic for estimating the difference in the means.

In this article we discuss the performance of these new confidence intervals. This article is structured as follows. In Section 2 we present the new methods for determination of one-sided nonparametric confidence intervals for the population variance and discuss the transformation and the bootstrap method. In Section 3 we compare the coverage accuracy of these methods with coverage accuracy of the χ^2 confidence intervals. In Section 4 we apply our methods to a real data set, composed of revenues of food retail stores, and stand out advantages of the transformed t statistic. In Section 5 we draw concluding remarks.

2. One-sided confidence interval for the variance

For interval estimation of the variance in case of one sample the dominant existing method is based on the χ^2 statistic. However, for enough large n, χ^2 distribution, with $n-1$ degrees of freedom, can be approximated by normal distribution with parameters $n-1$ and $2(n-1)$. If $X_1,\ldots,X_n$ is a random sample from distribution F with mean μ and variance σ^2 , and if $S^2=(1/(n-1))\sum_{i=1}^{n}(X_i-\overline{X})^2$ is a sample variance, then the distribution of the standardized variable

$$Z=\frac{\frac{(n-1)S^2}{\sigma^2}-(n-1)}{\sqrt{2(n-1)}}=\frac{S^2-\sigma^2}{\sqrt{\mathrm{var}(S^2)}}, \tag{3}$$

converges to standardized normal distribution as n increases to infinity (for details see Cojbasic and Tomovic, 2007). We may construct one-sided bootstrap-t confidence interval for σ^2 based on the t statistic

$$T=\frac{S^2-\sigma^2}{\sqrt{\widehat{\mathrm{var}}(S^2)}}, \tag{4}$$

where $\widehat{\mathrm{var}}(S^2)$ is a consistent estimator of the variance of S^2.

In each bootstrap sample (out of B bootstrap samples) we compute the value of statistic

$$T^*=\frac{S^{2*}-S^2}{\sqrt{\widehat{\mathrm{var}}(S^{2*})}}, \tag{5}$$

where S^{2*} is a bootstrap replication of statistic S^2. The upper endpoint and lower endpoint $1-\alpha$ level confidence intervals for σ^2 are intervals

$$I_{up}=\left(0,S^2-\widehat{t}^{(\alpha)}\sqrt{\widehat{\mathrm{var}}(S^2)}\right), \tag{6}$$

$$I_{low}=\left(S^2-\widehat{t}^{(1-\alpha)}\cdot\sqrt{\widehat{\mathrm{var}}(S^2)},+\infty\right), \tag{7}$$

where $\widehat{t}^{(1-\alpha)}$ and $\widehat{t}^{(\alpha)}$ are $1-\alpha$ and α percentiles of T^*.

If Cramer's condition holds (i.e. if $\limsup_{|t_1|+|t_2|\to\infty}|E\exp(it_1X_j+it_2X_j^2)|<1$) and if $E(X_j^6)<\infty$, T admits a first-order Edgeworth expansion

$$P(T\le x)=\Phi(x)+\frac{1}{\sqrt{n}}q(x)\phi(x)+O(n^{-1}), \tag{8}$$

where $q(x)=\frac{M'_3}{6}(2x^2+1)$, $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and density function, respectively. $M'_3=E((1/n)\sum_{i=1}^{n}X'^3_i)$, with random variables $X'_i, i=1,2,\ldots,n$, being defined in the following way:

$$X'_i=\frac{(X_i-\overline{X})^2-((n-1)/n)\sigma^2}{\sqrt{V_1}}\quad\text{for}\quad i=1,2,\ldots,n, \tag{9}$$

and $V_1=E((X_i-\overline{X})^2-((n-1)/n)\sigma^2)^2$ (for proof see Cojbasic and Tomovic, 2007). The Edgeworth expansion of Student's statistic was researched by several authors (Wallace, 1958; Bowman et al., 1977; Cressie, 1980; Singh, 1981; Hall, 1987, 1992b, etc.). Based on the Edgeworth expansion it is possible to suggest some transformation method aimed at removing the effect of skewness and getting the best coverage accuracy.