Advances and Applications in Statistics
© 2017 Pushpa Publishing House, Allahabad, India
http://www.pphmj.com
http://dx.doi.org/10.17654/AS051020151
Volume 51, Number 2, 2017, Pages 151-163 **ISSN: 0972-3617**

GENERALIZED CONFIDENCE INTERVALS FOR FUNCTION OF VARIANCES OF LOGNORMAL DISTRIBUTIONS

Suparat Niwitpong

Department of Applied Statistics
Faculty of Applied Science
King Mongkut's University of Technology North Bangkok
Thailand 10800

Abstract

This paper presents the generalized confidence intervals, proposed by Weerahandi [14], for a single variance, the difference between two population variances and the ratio of population variances of lognormal distribution. Monte Carlo simulation is used to evaluate the coverage probability of the proposed generalized confidence intervals. A real example is illustrated.

1. Introduction

Statistical inference for the function of population variances has been widely discussed by various authors, see for example, Arcos et al. [2], Singh et al. [11], Garcia and Cebrain [6] and Agrawal and Sthapit [1]. Kadilar and Cingi [7] proposed some ratio estimators for the population variance in simple and stratified random sampling. Cojbasic et al. [3] proposed the new

Received: May 18, 2017; Revised: June 14, 2017; Accepted: June 15, 2017
2010 Mathematics Subject Classification: 62F25.
Keywords and phrases: coverage probability, average length, variances.

method for the testing one population variance and the difference of variances of two samples, based on *t*-statistics and bootstrap method. Cojbasic and Tomovic [5] proposed the bootstrap methods to construct the confidence intervals of the population variance for one sample and the difference of variances of two samples. Cojbasic and Loncar [4] proposed one-sided of the bootstrap method to construct the confidence intervals of the population variance of skewed distributions. Singh and Malik [10] proposed a family of estimators for the population variance using auxiliary attributes. Phonyiem and Niwitpong [9] proposed the generalized confidence interval for the difference between normal variances. Somkhuen et al. [12] presented the upper bound of the generalized *p*-values for testing the variances of lognormal distributions. In this paper, we use the generalized confidence interval of Weerahandi [14], to construct confidence intervals for one population variance, the difference between two population variances and the ratio of population variances of lognormal distributions. The concepts of the generalized confidence interval introduced by Weerahandi [14] appear to be appropriate for the above problem, which involve developing confidence intervals in situations where traditional frequentist approaches do not provide useful solutions. In Section 2, the outlined for some basic steps to construct the generalized confidence intervals for this problem is presented. The process of simulation design for each of generalized confidence interval is presented in Section 3. Simulation results are shown in Section 4 and the application and the conclusion are presented in Sections 5-6.

2. Generalized Confidence Intervals for Function of Variances of Lognormal Distributions

Let $X^* = (X_1^*, X_2^*, ..., X_n^*)$ and $Y^* = (Y_1^*, Y_2^*, ..., Y_m^*)$ be random samples from two independent lognormal distributions with means μ_x, μ_y and standard deviations $\sigma_x, \sigma_y,$ respectively. It is well known that $\ln(X^*) = X \sim N(\mu_x, \sigma_x^2)$ and $\ln(Y^*) = Y \sim N(\mu_y, \sigma_y^2).$ The sample means and variances for X and Y are, respectively, denoted as $\bar{X}$, $\bar{Y}$, S_x^2 and S_y^2 when $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$, $\bar{Y} = n^{-1}\sum_{i=1}^{m} Y_i$, $S_x^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $S_y^2 = (m-1)^{-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$. Define also s_x^2 and s_y^2 are, respectively, sample variances from observed data of X and Y. We are interested in $100(1-\alpha)\%$ confidence intervals for $\theta_1 = \exp(2\mu_x + \sigma_x^2)(\exp(\sigma_x^2) - 1) = k_x$, $\theta_2 = k_x - k_y$, $k_y = \exp(2\mu_x + \sigma_x^2)(\exp(\sigma_x^2) - 1)$ and $\theta_3 = k_x / k_y$.

2.1. Generalized confidence interval for θ_1

It is not easy to find pivotal statistics for θ_1, θ_2 and θ_3. Hence, the generalized confidence intervals (GCI) proposed by Weerahandi [14] are recommended. We now give a brief introduction to the GCI idea based on Weerahandi [14]. The method of GCI is easy to use based on computational approach. To introduce the concept:

(A1) Let X and Y be random variables with probability distribution $f(X, Y, \theta, \varsigma)$, where θ is the parameter of interest and ς is a set of nuisance parameters.

(A2) Let (x, y) denote the observed value of (X, Y). To obtain a generalized confidence interval for θ, we start from the generalized pivotal quantity $T(X, Y; x, y, \theta, \varsigma)$, which is a function of the random variable (X, Y), its observed value (x, y) and the parameters θ and ς.

(A3) Also, $T(X, Y; x, y, \theta, \varsigma)$ is required to satisfy the following conditions:

(A3.1) For a fixed (x, y) the probability distribution of $T(X, Y; x, y, \theta, \varsigma)$ is free of unknown parameters.

(A3.2) The observed value of $T(X, Y; x, y, \theta, \varsigma)$, namely $T(x, y; x, y, \theta, \varsigma)$ is simply θ.

To obtain GCI for $\theta_1 = \exp(2\mu_x + \sigma_x^2)(\exp(\sigma_x^2) - 1)$, consider

$$\mu_x \approx \bar{x} - \frac{(\bar{X} - \mu_x)}{\sigma_x/\sqrt{n}} \frac{\sigma/\sqrt{n} s_x}{S_x} = \bar{x} - \frac{Z}{V/\sqrt{n-1}} \frac{s_x}{\sqrt{n}}$$

when $Z \sim N(0, 1)$, $V = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, χ^2_{n-1} is a Chi-square distribution with $n-1$ degrees of freedom and $\sigma_x^2 = \dfrac{(n-1)s_x^2}{V^2}$. Putting μ_x and σ_x^2 into $T_1(X, x, \theta_1)$, where

$$T_1(X, x, \theta_1)$$

$$= \exp\left(2\left(\bar{x} - \frac{Z}{V/\sqrt{n-1}} \frac{s}{\sqrt{n}}\right) + \frac{(n-1)s^2}{V^2}\right)\left(\exp\left(\frac{(n-1)s^2}{V^2}\right) - 1\right),$$

$T_1(X, x, \theta_1)$ is called a *generalized pivot statistic* for $\theta_1 = \exp(2\mu_x + \sigma_x^2)(\exp(\sigma_x^2) - 1)$ and it is easy to see that $T_1(X, x, \theta_1)$ is satisfied A3.1-A3.2. Hence, a $100(1-\alpha)\%$ generalized confidence interval for θ_1 is given by

$$CI_1 = [T_1(\alpha/2), T_1(1-\alpha/2)],$$

where α is a significance level at 0.05.

Algorithm A. To find the coverage probability for CI_1

1. For a set of data $X_1, ..., X_n$.

2. Compute $s_x^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$.

3. For *i* in 1 to *M*.

4. Generate $V = \chi^2_{n-1}$.

5. Set

$$T_1(X, x, \theta_1)$$

$$= \exp\left(2\left(\bar{x} - \frac{Z}{V/\sqrt{n-1}}\frac{s}{\sqrt{n}}\right) + \frac{(n-1)s^2}{V^2}\right)\left(\exp\left(\frac{(n-1)s^2}{V^2}\right) - 1\right).$$

6. End loop.

7. Set $A_i = 1$ if $T_{1(\alpha/2)} < \theta_1 < T_{1(1-\alpha/2)}$ else $A_i = 0$.

8. Coverage probability for confidence interval CI_1 is $M^{-1}\sum_{i=1}^{M} A_i$.

2.2. Generalized confidence interval for θ_2

Similarly to Subsection 2.1, the GCI for θ_2 is considered.

Let

$$\mu_y \approx \bar{y} - \frac{(\bar{Y} - \mu_y)}{\sigma_y/\sqrt{m}}\frac{\sigma_y/\sqrt{m}s_y}{S_y} = \bar{y} - \frac{Z}{U/\sqrt{m-1}}\frac{s_y}{\sqrt{m}}$$

and

$$U = \frac{(m-1)S_y^2}{\sigma_y^2} \sim \chi_{m-1}^2,$$

χ_{m-1}^2 is a Chi-square distribution with $m-1$ degrees of freedom and $\sigma_y^2 = \frac{(m-1)s_y^2}{U^2}$. Putting μ_x, μ_y and σ_x^2, σ_y^2 into $T_2(X, x, Y, y, \theta_2)$, where

$$T_2(X, x, Y, y, \theta_2)$$

$$= \exp\left(2\left(\bar{x} - \frac{Z}{V/\sqrt{n-1}}\frac{s_x}{\sqrt{n}}\right) + \frac{(n-1)s_x^2}{V^2}\right)\left(\exp\left(\frac{(n-1)s_x^2}{V^2}\right) - 1\right)$$

$$- \exp\left(2\left(\bar{y} - \frac{Z}{U/\sqrt{m-1}}\frac{s_y}{\sqrt{m}}\right) + \frac{(m-1)s_y^2}{U^2}\right)\left(\exp\left(\frac{(m-1)s_y^2}{U^2}\right) - 1\right).$$

$T_2(X, x, Y, y, \theta_2)$ is called a *generalized pivot statistic* for

$$\theta_2 = \exp\left(2\left(\bar{x} - \frac{Z}{V/\sqrt{n-1}}\frac{s_x}{\sqrt{n}}\right) + \frac{(n-1)s_x^2}{V^2}\right)\left(\exp\left(\frac{(n-1)s_x^2}{V^2}\right) - 1\right)$$

$$- \exp\left(2\left(\bar{y} - \frac{Z}{U/\sqrt{m-1}}\frac{s_y}{\sqrt{m}}\right) + \frac{(m-1)s_y^2}{U^2}\right)\left(\exp\left(\frac{(m-1)s_y^2}{U^2}\right) - 1\right)$$

and it is easy to see that $T_2(X, x, Y, y, \theta_2)$ is satisfied A3.1-A3.2. Hence, a $100(1-\alpha)\%$ generalized confidence interval for θ_2 is given by $CI_2 =$ $[T_2(\alpha/2), T_2(1-\alpha/2)]$.

Algorithm B. To find the coverage probability for CI_2 is similarly to Algorithm A.

2.3. Generalized confidence interval for θ_3

Similarly to Subsections 2.1 and 2.2, the GCI for θ_3 is given by

$$T_3(X, x, Y, y, \theta_3)$$

$$= \frac{\exp\left(2\left(\bar{x} - \frac{Z}{V/\sqrt{n-1}}\frac{s_x}{\sqrt{n}}\right) + \frac{(n-1)s_x^2}{V^2}\right)\left(\exp\left(\frac{(n-1)s_x^2}{V^2}\right) - 1\right)}{\exp\left(2\left(\bar{y} - \frac{Z}{U/\sqrt{m-1}}\frac{s_y}{\sqrt{m}}\right) + \frac{(m-1)s_y^2}{U^2}\right)\left(\exp\left(\frac{(m-1)s_y^2}{U^2}\right) - 1\right)}.$$

It is also easy to see that $T_3(X, x, Y, y, \theta_3)$ is satisfied A3.1-A3.2. Hence, a $100(1-\alpha)\%$ generalized confidence interval for θ_3 is given by $CI_3 =$ $[T_3(\alpha/2), T_3(1-\alpha/2)]$.

Algorithm C. To find the coverage probability for CI_3 is similarly to Algorithm A.

In the next section, we evaluate these confidence intervals based on their coverage probabilities. Typically, we prefer a confidence interval whose coverage probability is at least or close to the nominal level $1-\alpha = 0.95$. Monte Carlo simulation will be used to assess this criterion.

3. Simulation Framework

In this section, we use Monte Carlo simulation to assess three generalized confidence intervals notified in the previous section: CI_1, CI_2, CI_3 based on their coverage probabilities. We design a simulation, without losing generality, by setting $\mu_x = 1$ and $\sigma_x =$ 0.2, 0.5, 0.7, 1, 1.3, 1.5, 2, 3, $n =$ 15, 30, 50, 100, 200 for CI_1. Also setting $\mu_1 = \mu_2 = 1$, a ratio of variances $\sigma_x/\sigma_y =$ 0.2, 0.5, 0.7, 1, 1.3, 1.5, 2, 3 and the sample sizes $(n, m) = (15, 30)$, $(30, 15)$, $(30, 30)$, $(30, 50)$, $(50, 30)$, $(50, 50)$, $(50, 100)$, $(100, 50)$, $(100, 100)$, $(100, 200)$, $(200, 100)$, $(200, 200)$. We wrote function in R program [13] to generate the data with lognormally distributed with means and variances which are mentioned previously to construct two confidence intervals, i.e., CI_1, CI_2, CI_3 and then compute coverage probability of each confidence interval. All results are illustrated in Tables 1 and 2 with a number of simulation runs, $M = 5000$, GCI = 2000 and nominal level $(1-\alpha = 0.95)$.

Table 1. The coverage probabilities of 95% of two-sided confidence intervals for variance of lognormal distribution (based on 5000 simulations) $\mu = 1$, GCI = 2000

n	σ_x	CI_1
15	**0.2**	0.9486
	0.5	0.9538
	0.7	0.9552
	1.0	0.9472
	1.3	0.9438
	1.5	0.9494
	2.0	0.9466
	3.0	0.9538

30	**0.2**	0.9506
	0.5	0.9500
	0.7	0.9460
	1.0	0.9522
	1.3	0.9496
	1.5	0.9530
	2.0	0.9518
	3.0	0.9456
50	**0.2**	0.9464
	0.5	0.9488
	0.7	0.9544
	1.0	0.9544
	1.3	0.9506
	1.5	0.9482
	2.0	0.9436
	3.0	0.9458
100	**0.2**	0.9542
	0.5	0.9520
	0.7	0.9552
	1.0	0.9518
	1.3	0.9516
	1.5	0.9506
	2.0	0.9484
	3.0	0.9490
200	**0.2**	0.9564
	0.5	0.9502
	0.7	0.9522
	1.0	0.9502
	1.3	0.9502
	1.5	0.9488
	2.0	0.9440
	3.0	0.9522

Table 2. The coverage probabilities of 95% of two-sided confidence intervals for the difference between variances and the ratio of variances of lognormal distribution (based on 5000 simulations) $\mu_1 = \mu_2 = \mu = 1$, $\text{GCI} = 2000$

N	m	σ_x/σ_y	CI_2	CI_3
15	**30**	**0.2**	0.9564	0.9522
		0.5	0.9496	0.9516
		0.7	0.9536	0.9508
		1.0	0.9444	0.9444
		1.3	0.9524	0.9570
		1.5	0.9464	0.9496
		2.0	0.9466	0.9494
		3.0	0.9538	0.9564
30	**15**	**0.2**	0.9534	0.9546
		0.5	0.9464	0.9476
		0.7	0.9520	0.9492
		1.0	0.9516	0.9516
		1.3	0.9574	0.9498
		1.5	0.9604	0.9548
		2.0	0.9522	0.9494
		3.0	0.9456	0.9498
30	**30**	**0.2**	0.9534	0.9534
		0.5	0.9450	0.9500
		0.7	0.9540	0.9504
		1.0	0.9540	0.9542
		1.3	0.9532	0.9498
		1.5	0.9522	0.9520
		2.0	0.9510	0.9468
		3.0	0.9488	0.9474
30	**50**	**0.2**	0.9502	0.9526
		0.5	0.9480	0.9488
		0.7	0.9556	0.9494
		1.0	0.9458	0.9456
		1.3	0.9446	0.9472
		1.5	0.9492	0.9488
		2.0	0.9492	0.9512
		3.0	0.9470	0.9484

50	**30**	**0.2**	0.9512	0.9510
		0.5	0.9498	0.9516
		0.7	0.9502	0.9528
		1.0	0.9544	0.9544
		1.3	0.9552	0.9522
		1.5	0.9514	0.9492
		2.0	0.9436	0.9464
		3.0	0.9458	0.9476
50	**50**	**0.2**	0.9446	0.9478
		0.5	0.9472	0.9496
		0.7	0.9504	0.9440
		1.0	0.9468	0.9468
		1.3	0.9554	0.9484
		1.5	0.9518	0.9494
		2.0	0.9482	0.9484
		3.0	0.9480	0.9478
50	**100**	**0.2**	0.9530	0.9514
		0.5	0.9492	0.9516
		0.7	0.9586	0.9526
		1.0	0.9414	0.9414
		1.3	0.9530	0.9494
		1.5	0.9558	0.9526
		2.0	0.9490	0.9470
		3.0	0.9534	0.9538
100	**50**	**0.2**	0.9510	0.9496
		0.5	0.9506	0.9468
		0.7	0.9486	0.9500
		1.0	0.9538	0.9538
		1.3	0.9534	0.9508
		1.5	0.9504	0.9514
		2.0	0.9482	0.9490
		3.0	0.9490	0.9502
100	**100**	**0.2**	0.9482	0.9456
		0.5	0.9528	0.9488
		0.7	0.9486	0.9460
		1.0	0.9532	0.9532
		1.3	0.9476	0.9516
		1.5	0.9502	0.9510
		2.0	0.9474	0.9472

		3.0	0.9482	0.9492
100	**200**	**0.2**	0.9458	0.9452
		0.5	0.9484	0.9502
		0.7	0.9514	0.9510
		1.0	0.9508	0.9508
		1.3	0.9518	0.9520
		1.5	0.9554	0.9520
		2.0	0.9458	0.9438
		3.0	0.9508	0.9496
200	**100**	**0.2**	0.9484	0.9472
		0.5	0.9516	0.9568
		0.7	0.9522	0.9532
		1.0	0.9472	0.9472
		1.3	0.9500	0.9528
		1.5	0.9482	0.9496
		2.0	0.9444	0.9468
		3.0	0.9522	0.9528
200	**200**	**0.2**	0.9500	0.9510
		0.5	0.9444	0.9466
		0.7	0.9516	0.9478
		1.0	0.9530	0.9530
		1.3	0.9466	0.9412
		1.5	0.9484	0.9490
		2.0	0.9548	0.9512
		3.0	0.9492	0.9480

4. Simulation Results

From Tables 1 and 2, we found that, for every case, coverage probabilities of confidence intervals CI_1, CI_2, CI_3 are generally closed to or above the nominal level 0.95 for all sample sizes. As a result, the generalized confidence interval is recommended for constructing the confidence interval for the function of variances of lognormal distribution.

5. Application

The data is taken from Table 1 of Lee and Lin [8] to measure the relative level of carboxyhemoglobin for a group of nonsmokers and a group of cigarette smokers. The confidence interval for CI_1 is (136.2581, 1135.3126) with length equals 999.0545. The confidence interval for CI_2 is $(-469414079, -1084933)$ with length equals 68329146 and the confidence interval for CI_3 is (0.0000, 0.0004) with length equals 0.0004. The results from confidence intervals CI_2 and CI_3 explain that a measurement of the relative level of carboxyhemoglobin for a group of nonsmokers and a group of cigarette smokers is different.

6. Conclusion

In this paper, we proposed new confidence intervals for the single variance, the difference variances and the ratio of variances of lognormal distribution based on Weerahandi [14]. The GCI perform well in terms of coverage probabilities. For almost cases, we found that coverage probabilities of the intervals, CI_1, CI_2, CI_3 for all sample sizes and ratios of variances, are very closed to (or above) the nominal level 0.95. We therefore recommend the GCI for constructing the confidence interval for the function of variances of lognormal distribution.

Acknowledgement

This research was funded by King Mongkut's University of Technology North Bangkok. Contract no. KMUTNB-60-GOV-062.

The author thanks the anonymous referees for their valuable suggestions which led to the improvement of the manuscript.

References

[1] M. C. Agrawal and A. B. Sthapit, Unbiased ratio-type variance estimation, Statist. Probab. Lett. 25 (1995), 361-364.

[2] A. Arcos, M. Rueda, M. D. Martinez, S. Gonzalez and Y. Roman, Incorporating the auxiliary information available in variance estimation, Appl. Math. Comput. 160 (2005), 387-399.

[3] V. Cojbasic, J. Kocovic, D. Loncar and T. R. Antic, Testing population variance in case of one sample and the difference of variances in case of two samples: Example of wage and pension data sets in Serbia, Economic Modelling 29 (2012), 610-613.

[4] V. Cojbasic and D. Loncar, One-sided confidence intervals for population variances of skewed distributions, J. Statist. Plann. Inference 141 (2011), 1667-1672.

[5] V. Cojbasic and A. Tomovic, Nonparametric confidence intervals for population variance of one sample and the difference of variances of two samples, Comput. Statist. Data Anal. 51 (2007), 5562-5578.

[6] M. R. Garcia and A. A. Cebrain, Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator, Metrika 45 (1997), 171-178.

[7] C. Kadilar and H. Cingi, Ratio estimators for the population variance in simple and stratified random sampling, Appl. Math. Comput. 173 (2006), 1047-1059.

[8] J. C. Lee and S.-H. Lin, Generalized confidence intervals for the ratio of means of two normal populations, J. Statist. Plann. Inference 123 (2004), 49-60.

[9] W. Phonyiem and S. A. Niwitpong, Generalized confidence interval for the difference between normal population variances, Far East J. Math. Sci. (FJMS) 69(1) (2012), 99-110.

[10] R. Singh and S. Malik, Improved estimation of population variance using information on auxiliary attribute in simple random sampling, Appl. Math. Comput. 235 (2014), 43-49.

[11] H. P. Singh, L. N. Upadhyaya and U. D. Namjoshi, Estimation of finite population variance, Current Science 57 (1988), 1331-1334.

[12] R. Somkhuen, S. Niwitpong and S. A. Niwitpong, Upper bound of the generalized *p*-value for the population variances of lognormal distribution with known coefficients of variation, J. Probab. Statist. (2017), Article ID 4793702, pp. 1-9.

[13] The R Development Core Team, An Introduction to R, Vienna: R Foundation for Statistical Computing, http://www.R-project.org, 2010.

[14] S. Weerahandi, Generalized confidence intervals, J. Amer. Statist. Assoc. 88 (1993), 899-905.