Thailand Statistician
January 2008; 6(1) : 1-6
http://statassoc.or.th
Contributed paper

Adjusted Bonett Confidence Interval for Standard Deviation of Non-normal Distributions

Sa-aat Niwitpong* and Pianpool Kirdwichai
Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's Institute of Technology North Bangkok, Bangkok 10800, Thailand.
***Author for correspondence**; e-mail: snw@kmitnb.ac.th

Received: 20 June 2007
Accepted: 19 November 2007.

Abstract

Bonett [1] provides an approximate confidence interval for σ and shows it to be nearly exact under normality and, for small samples, under moderate non-normality. This paper proposes a new method for determining the confidence interval for σ. We follow the suggestion of Bonett [1] and include any prior kurtosis information. An important modification from the Bonett method is to base the interval on $t_{\alpha/2}$, the $(\alpha/2)100th$ quantile of Student T distribution, rather than on $Z_{\alpha/2}$, the $(\alpha/2)100th$ quantile of standard normal distribution. Further, the use of the median as an estimate of the population mean gives a slightly higher coverage probability for the confidence interval for σ when data are from skewed leptokurtic distributions. Monte Carlo Simulation results for selected normal and non-normal distributions show that the confidence intervals obtained from the new method have appreciably higher coverage probabilities than the confidence intervals from the original Bonett method that does not use prior kurtosis information, and also higher coverage probabilities than the Bonett method that does use prior kurtosis information.

Keyword: confidence interval, coverage probability, standard deviation.

1. Introduction

Let $X_1, X_2, ..., X_n$ be a random sample of size n. If $X_i \sim N(\mu, \sigma^2)$ then Bonett [1] has stated that an exact *100(1-α)%* confidence interval for σ^2 is:

$$(n-1)S^2/\chi^2_{1-\alpha/2} < \sigma^2 < (n-1)S^2/\chi^2_{\alpha/2} \quad (1)$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are the *(α/2)100*th and *(1-α/2)100*th percentiles of the central chi-square distribution with $\upsilon = n-1$ degrees of freedom and $\overline{X}$ and s^2 are the well-known sample mean and sample variance respectively. Taking the square root of the endpoints of (1) gives a confidence interval for σ.

The exact confidence interval (1) is sensitive to violation of the normality assumption. For some symmetric and skewed leptokurtic distributions, simulation studies of Bonett [1] show that, if α = 0.05 and n = 25, then (1) gives asymptotic coverage probabilities of about 0.833, 0.722, and 0.594 for the t(5), exponential(2), and $\chi^2(1)$, distributions respectively. These coverage probabilities are not close to a nominal value 0.95.

For non-normal distributions, Bonett [1] suggested the following modifications to the normal distribution formula given in (1). Bonett [1] used the results that the variance of s^2 may be expressed as $\sigma^4\{\gamma_4-(n-3)/(n-1)\}/n$ where $\gamma_4=\mu^4/\sigma^4$ is the kurtosis of the distribution and μ^4 is the population fourth central moment (Mood et al. , [3], pp. 229). Bonett [1] then combined the properties of *ln* s^2, whose sampling distribution converges to normality faster than the sampling distribution of s^2 when $X_i \sim N(\mu,\sigma^2)$, with the formula for the small-sample adjustment ratio, $\{\gamma_4 - (n-3)/n\}/(n-1)$, found by Shoemaker [7]. Bonett then obtained the following *100(1-α)%* confidence interval for σ^2

$$\text{Exp}\{\ln(\text{cS}^2) - Z_{\alpha/2}Se\} < \sigma^2 < \text{Exp}\{\ln(\text{cS}^2) + Z_{\alpha/2}Se\} \quad (2)$$

where $Z_{\alpha/2}$ is the *(1-α/2) 100*th percentile of the standard normal distribution,

$$Se = c[\{\hat{\gamma}_4^*(n-3)/n\}/(n-1)]^{1/2}, \quad (3)$$

$c = n/(n- Z_{\alpha/2})$ is a small-sample adjustment, $\hat{\gamma}_4^* = (n_0\tilde{\gamma}_4 + n\bar{\gamma}_4)/(n_0 + n)$, $\tilde{\gamma}_4$ is a prior estimate of γ_4 obtained from a larger sample of size n_0, and

$$\bar{\gamma}_4 = \frac{n\sum_{i=1}^{n}(X_i - \bar{X})^4}{(\sum_{i=1}^{n}(X_i - \bar{X})^2)^2} \qquad (4)$$

where m is a trimmed mean with trim-proportion equal to $1/\{2(n-4)^{1/2}\}$. Bonett also suggested combining the estimate $\bar{\gamma}_4$ of γ_4 obtained from the small sample of size n with $\tilde{\gamma}_4$ to give a pooled estimate of γ_4 as $\hat{\gamma}_4^* = (n_0\tilde{\gamma}_4 + n\bar{\gamma}_4)/(n_0 + n)$. However, if $\tilde{\gamma}_4$ is not available then $\hat{\gamma}_4^* = \bar{\gamma}_4$ should be used.

Our aim in this paper is to replace the Bonett confidence interval for σ with a confidence interval that has a higher coverage probability for non-normal data. Pan [6] and Olsson [5] found that, for small sample sizes, replacing $Z_{\alpha/2}$ with $t_{\alpha/2}$, gave higher coverage probabilities of the confidence intervals for σ for binomial distributions for a range of proportions and for the mean for Log-Normal distribution respectively. It is hoped to improve the coverage probability of the Bonett formula by replacing $Z_{\alpha/2}$ with $t_{\alpha/2}$ and by replacing the trimmed mean with the median, which is a good estimator of μ when data are from asymmetric and skewed leptokurtic distributions.

In section 2 an adjusted confidence interval for σ is described. In section 3, simulation results of coverage probabilities of confidence intervals are given for selected non-normal distributions with a range of values of kurtosis. Section 4 contains a discussion of the results and conclusions.

2. Proposed Confidence Interval

From (2) and (4), replacement of $Z_{\alpha/2}$ with $t_{\alpha/2}$ gives a proposed confidence interval for σ of

$$\text{Exp}\{\ln \text{cS}^2 - t_{\alpha/2,\text{v}}Se\} < \sigma^2 < \text{Exp}\{\ln \text{cS}^2 + t_{\alpha/2,\text{v}}Se\} \qquad (5)$$

where $t_{\alpha/2}$ is the $(1-\alpha/2)\ 100^{th}$ percentile of the t-distribution with $\upsilon = n-1$ degrees of freedom and $c = n/(n- t_{\alpha/2})$ is as stated in (3).

In (5), we modify the definition of *Se* given in (3) as follows. For non-normal distributions, we propose the median as an estimator of μ. On substituting the trimmed mean *m* in (4) with a median value, we obtain the estimator of γ_4 as

$$\gamma_4^{/} = \frac{n\sum_{i=1}^{n}(X_i - med)^4}{(\sum_{i=1}^{n}(X_i - \overline{X})^2)^2} \qquad (6)$$

where *med* is the median. A pooled estimator of γ_4 that uses both prior and sample kurtosis information is given by:

$$\gamma_4^{*} = \frac{n_0\widetilde{\gamma}_4 + n\gamma_4^{/}}{n_0 + n} \qquad (7)$$

3. Simulation Results

Using 50,000 Monte Carlo simulation samples from programs written in R, we obtained the estimated coverage probabilities for σ for the normal distribution formula (1), the Bonett formula with and without a prior kurtosis estimate (2), and our formula also with and without prior kurtosis (5). In our formula the estimated values of $\hat{\gamma}_4$ were obtained either from (4) or the pooled estimator (7), and the sample estimates of μ were used either the trimmed mean or the median. The simulation results are reported in Tables 1 and 2.

We chose α = 0.05. It can be seen from Table 1 that the new formula performs better than the Bonett formula for all of the six distributions studied. A comparison of the coverage probabilities in Tables 1 and 2 shows that for non-normal distributions the formula with kurtosis information appreciably improves the coverage probabilities.

4. Discussion and Conclusions

In this paper, we have presented a comparison between coverage probabilities for the confidence intervals for σ obtained from the Bonett confidence interval (2) and our modified Bonett confidence interval (5). Kabaila [2] and Niwitpong [4] have argued that confidence intervals can only be evaluated solely on the basis of their expected length when each of confidence interval has a minimum coverage probability $1-\alpha$ where α is a significance level. In this paper, some of confidence intervals (2) and (5) have coverage probabilities below the nominal level $1-\alpha$, (say 0.95). We have therefore compared only the coverage probabilities and not the lengths of the confidence intervals (2) and (5).

Table 1 : Estimated 95% coverage probabilities of (1) , (2) and (5) for six non-normal distributions.

Distribution	***n***	**(1)**	**(2)**	**(5)**	Distribution	***n***	**(1)**	**(2)**	**(5)**
Uniform	10	0.993	0.970	0.996	*t(5)*	10	0.874	0.938	0.995
	25	0.997	0.950	0.996		25	0.833	0.912	0.987
	50	0.997	0.949	0.996		50	0.798	0.908	0.975
	100	0.997	0.948	0.997		100	0.784	0.914	0.967
Beta(3,3)	10	0.979	0.965	0.992	*Exp*	10	0.766	0.888	0.993
	25	0.981	0.951	0.990		25	0.722	0.890	0.980
	50	0.981	0.949	0.989		50	0.697	0.899	0.972
	100	0.982	0.950	0.989		100	0.685	0.917	0.957
Beta(1,10)	10	0.829	0.912	0.952	$\chi^2(1)$	10	0.640	0.850	0.984
	25	0.805	0.912	0.940		25	0.594	0.860	0.963
	50	0.798	0.925	0.948		50	0.565	0.880	0.950
	100	0.793	0.935	0.959		100	0.562	0.900	0.944

Table 2 : Effect of prior kurtosis information on the 95% coverage probabilities of (5).

Distribution	n_0	$\tilde{\gamma}_4$	n	Bonett (2)	New formula (5)	
					$\hat{\mu}$ =m	$\hat{\mu}$ =med
Uniform	200	1.8	25	0.945	0.995	0.999
			100	0.948	0.997	0.997
	500	1.8	25	0.944	0.996	0.996
			100	0.949	0.999	0.998
t(5)	200	6.1	25	0.971	0.986	0.997
			100	0.948	0.966	0.978
	500	7.0	25	0.981	0.991	0.998
			100	0.962	0.974	0.979
Exp	200	7.9	25	0.959	0.965	0.965
			100	0.941	0.960	0.969
	500	8.5	25	0.968	0.985	0.993
			100	0.952	0.957	0.974
$\chi^2(1)$	200	12.2	25	0.950	0.960	0.961
			100	0.934	0.954	0.954
	500	13.6	25	0.964	0.975	0.975
			100	0.947	0.950	0.965

A modification for the Bonett formula for the confidence interval for σ is given by Bonett [1]. When prior kurtosis information is not used, this new confidence interval gives better coverage probabilities than the confidence interval of Bonett [1] (see Table 1). When prior kurtosis information is used, our new formula (5) also gives an improvement of the confidence interval, based on its coverage probability, compared to the confidence interval given by the Bonett formula (2) (see Table 2). In both cases the effect of replacing the population mean with the trimmed mean and the median has been examined. We found that, for some non-normal distributions, estimating the population mean with the median gives higher coverage probabilities than estimating the population mean with the trimmed mean in the confidence interval given in (5). The simulation results in both Tables 1 and 2 show that the new confidence interval for σ is wider than the confidence interval given in (2) especially for the cases of small sample sizes, i.e., n= 10 and 25.

5. Acknowledgements

The authors wish to thank three anonymous referees and Dr. Elvin Moore for helpful comments.

References

[1] Bonett, D. Approximate confidence interval for standard deviation of nonnormal distributions. *Comput. Statist. Data Anal*. 2006; 50: 775-782.

[2] Kabaila, P. Valid confidence intervals in regression after variable selection. Econometric Theory. 14; 1998: 463-482.

[3] Mood, A. M., Graybill, F. A. Boes, D. C. *Introduction to the Theory of Statistics, third ed.* New York: McGraw-Hill. 1974.

[4] Niwitpong, S. *Predictive inference for times series*. Ph.D thesis (submitted for examination), Department of Statistical Science, La Trobe University, 2005.

[5] Olsson, U. Confidence Intervals for the Mean of a Log-Normal Distribution. *J. Statist. Educat*. 2005; 13, available from (http://www.amstat.org/publications/jse/v13n1/olsson.html).

[6] Pan, W. Approximate confidence intervals for one proportion and difference of two proportions. *Comput. Statist. Data Anal*. 2002; 40: 143-157.

[7] Shoemaker, L.H. Fixing the F test for equal variances. *The Amer. Statisticians*. 2003; 57: 105-114.