Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computational Statistics & Data Analysis 50 (2006) 775–782

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Approximate confidence interval for standard deviation of nonnormal distributions☆

Douglas G. Bonett*

Department of Statistics, Iowa State University, Snedechor Hall, Ames, IA 50011, USA

Received 10 May 2004; received in revised form 6 October 2004; accepted 7 October 2004
Available online 29 October 2004

Abstract

The exact confidence interval for σ is hypersensitive to minor violations of the normality assumption and its performance does not improve with increasing sample size. An approximate confidence interval for σ is proposed and is shown to be nearly exact under normality with excellent small-sample properties under moderate nonnormality. The small-sample performance of the proposed interval may be further improved using prior kurtosis information. A sample size planning formula is given.

Keywords: Dispersion; Interval estimation; Kurtosis; Prior information; Sample size; Variability

1. Introduction

Let $Y_1, Y_2, \ldots, Y_n$ be a random sample. If $Y_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$ for all i, then an exact $100(1-\alpha)\%$ confidence interval for σ^2 is

$$(n-1)\hat{\sigma}^2/U < \sigma^2 < (n-1)\hat{\sigma}^2/L, \tag{1}$$

where $U = \chi^2_{\alpha/2;n-1}$, $L = \chi^2_{1-\alpha/2;n-1}$, $\hat{\sigma}^2 = \sum\left(Y_i - \hat{\mu}\right)^2/(n-1)$, $\hat{\mu} = \sum Y_i/n$, $\chi^2_{p,\mathrm{df}}$ is the point on a central chi-squares distribution with df degrees of freedom exceeded with

☆ This research was supported by Grant SES-0343552 from the Methodology, Measurement, and Statistics program of the US National Science Foundation.

* Tel.: +5152941975; fax: +5152944040.
E-mail address: dgbonett@iastate.edu (D.G. Bonett).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2004.10.003

probability p (Tate and Klett, 1959). Taking the square root of the endpoints of (1) gives a confidence interval for σ.

The exact confidence interval (1) is hypersensitive to minor violations of the normality assumption. The results of Scheffé (1959, p. 336) can be applied to show that (1) has an asymptotic coverage probability of about 0.76, 0.63, 0.60, and 0.51 for the Logistic, $t(7)$, Laplace, and $t(5)$ distributions, respectively. This result is disturbing because these symmetric distributions are not easily distinguished from a normal distribution unless the sample size is large. Miller (1986, p. 264) describes this situation as "catastrophic".

An alternative to the exact confidence interval is proposed here that: (1) is nearly exact under normality, (2) has coverage probability close to $1-\alpha$ under moderately nonnormality, (3) has coverage probability that approaches $1-\alpha$ as the sample size increases for nonnormal distributions with finite fourth moments, and (4) is not computationally intensive.

2. Proposed confidence interval

Instead of assuming $Y_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$, let $Y_i \ (i = 1, 2, \ldots, n)$ be continuous, independent and identically distributed random variables with $0 < var\,(Y_i) = \sigma^2$, $E\,(Y_i) = \mu$ and finite fourth moment. The variance of $\hat{\sigma}^2$ may be expressed as $\sigma^4 \left\{\gamma_4 - (n-3)/(n-1)\right\}/n$, where $\gamma_4 = \mu^4/\sigma^4$ and μ^4 is the population fourth central moment (Mood et al., 1974, p. 229). A variance-stabilizing transformation for $\hat{\sigma}^2$ is $\ln\left(\hat{\sigma}^2\right)$ and application of the delta method gives $var\ \ln\left(\hat{\sigma}^2\right) \cong \left\{\gamma_4 - (n-3)/(n-1)\right\}/n$. Shoemaker, 2003 found that using $\left\{\gamma_4 - (n-3)/n\right\}/(n-1)$ improved the small-sample performance of his equal-variance test, and this small-sample adjustment will be used here. In practice, γ_4 is unknown and an estimate of $var\ \ln\left(\hat{\sigma}^2\right)$ will require an estimate of γ_4. Pearson's estimator $\hat{\gamma}_4 = n\sum\left(Y_i - \hat{\mu}\right)^4/\left(\sum\left(Y_i - \hat{\mu}\right)^2\right)^2$ tends to have large negative bias in leptokurtic (heavy tailed) distributions unless the sample size is very large. The following estimator of γ_4, which is asymptotically equivalent to Pearson's estimator, is proposed

$$\bar{\gamma}_4 = n\sum\left(Y_i - m\right)^4 \Big/ \left(\sum\left(Y_i - \hat{\mu}\right)^2\right)^2, \tag{2}$$

where m is a trimmed mean with trim-proportion equal to $1/\left\{2(n-4)^{1/2}\right\}$ so that m converges to μ as n increases without bound. This estimator of kurtosis tends to have less negative bias and smaller coefficient of variability than Pearson's estimator in symmetric and skewed leptokurtic distributions.

In some applications a large-sample estimate of γ_4 from a previous study will be available. Let $\tilde{\gamma}_4$ denote a prior point estimate of γ_4 obtained from a sample of size n_0. The prior point estimate may be combined with (2) to give a pooled estimate of γ_4

$$\hat{\gamma}_4^* = \left(n_0\tilde{\gamma}_4 + n\bar{\gamma}_4\right)/(n_0 + n), \tag{3}$$

which obviously simplifies to (2) when prior information is unavailable.

A prior point estimate of γ_4 need not come from a single large sample but instead could be a pooled estimate from several small samples. When pooling kurtosis estimates from several small samples, Laylard (1973) recommends pooling the fourth moments, pooling the variances, and then computing the kurtosis estimate from these pooled estimates. The fourth moment may be expressed as $\sigma^4\gamma_4$ so that prior estimates of σ^2 and γ_4 may be used to obtain a pooled estimate of γ_4.

In addition to the variance-stabilizing property of $\ln\left(\hat{\sigma}^2\right)$, Bartlett and Kendall (1946) show that the sampling distribution of $\ln\left(\hat{\sigma}^2\right)$ converges to normality much faster than the sampling distribution of $\hat{\sigma}^2$ when $Y_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$. Scheffé (1959, p. 84) and Laylard (1973) recommend the logarithmic transformation for nonnormal distributions as well. Given the desirable properties of $\ln\left(\hat{\sigma}^2\right)$, a large-sample confidence interval for σ^2 may be obtained from a reverse-transformed confidence interval for $\ln\left(\sigma^2\right)$. The following $100(1-\alpha)\%$ confidence interval for σ^2 is proposed

$$\exp\left\{\ln\left(c\hat{\sigma}^2\right) \pm z_{\alpha/2}se\right\}, \tag{4}$$

where $z_{\alpha/2}$ is two-sided critical z-value, $se = c\left[\left\{\hat{\gamma}_4^*(n-3)/n\right\}/(n-1)\right]^{1/2}$, and $c = n/\left(n - z_{\alpha/2}\right)$ is an empirically determined, small-sample adjustment that helps equalize the tail probabilities. Taking the square root of the endpoints of (4) gives a confidence interval for σ. Simulations suggest that when $n_0 > n$, replacing $(n-3)/n$ with 1 and replacing $n-1$ with n in se will improve the small-sample performance of (4).

3. Simulation results

Estimates of coverage probabilities and average interval widths of (1) and (4) were obtained using 50,000 Monte Carlo random samples of a given sample size from various distributions. The simulation programs were written in Gauss and executed on a Pentium 4 computer.

The performance of (4) for normal distributions is examined first. Estimated coverage probabilities of (4) and estimated average confidence intervals widths for both (1) and (4) are displayed in Table 1. Prior kurtosis information is not utilized in (4) for this simulation. The results in Table 1 suggest that (4) has coverage probability close to $1-\alpha$ when sampling from a normal distribution with $n > 10$. If (4) is used in a sample of size $n + 3$ then its average width will be about the same as the average width of (1) from a sample size of n. This is remarkable because Cohen (1972) has shown that no other confidence interval based on $\hat{\sigma}^2$ is shorter than (1). The cost of sampling three additional units reflects the cost of using (4) instead of (1) in those rare applications where the distribution is known with certainty to be closely approximated by a normal. This will be a small price to pay if (4) performs substantially better than (1) when sampling from nonnormal distributions. The next simulation compares the performance of (1) and (4) for a variety of nonnormal distributions.

Estimated coverage probabilities ($\alpha = 0.05$) of (1) and (4) for several nonnormal distributions are displayed in Table 2. Prior kurtosis information is not utilized in (4) for this simulation. The results in Table 2 suggest that (4) is slightly conservative in platykurtic

Table 1
Estimated coverage probabilities for (4) and average widths of (1) and (5) for normal distributions

$1-\alpha$	n	A	B	C
0.90	10	0.915	2.60	2.25
	25	0.897	1.66	1.62
	50	0.895	1.40	1.40
	100	0.896	1.27	1.26
0.95	10	0.960	3.28	2.65
	25	0.946	1.86	1.78
	50	0.945	1.50	1.49
	100	0.945	1.33	1.32
0.99	10	0.994	5.57	3.69
	25	0.987	2.29	2.15
	50	0.987	1.72	1.69
	100	0.988	1.45	1.45

Note: A = estimated coverage probability of (4); B = estimated average width of (4); C = estimated average width of (1).

distributions and slightly liberal in moderately leptokurtic or skewed distributions. As expected, the coverage probability of (4) improves as n increases. With highly nonnormal distributions the coverage probability of (4) can be considerably less than $1-\alpha$ unless n is large. In contrast to (4), (1) is very conservative in platykurtic distributions, very liberal in leptokurtic distributions, and its coverage probability does not converge to $1-\alpha$ as n increases. Clearly (4) is superior to (1) for all distributions considered in Table 2.

The results in Table 3 illustrate the effect of prior kurtosis information on the small-sample ($n=10$, 100) performance of (4). This simulation describes the effect of using a prior point estimate of Pearson's kurtosis from a sample of $n_0 = 200$ or 500. The prior point estimates shown in Table 3 are approximate expected values of Pearson's kurtosis estimator at n_0. The results in Table 3 suggest that using prior kurtosis estimates can improve the small-sample performance of (4) in most cases. Of course, prior kurtosis information would degrade the performance of (4) if $n_0 - n$ and $|\gamma_4 - \tilde{\gamma}_4|$ are both large. It should be noted that the performance of (4) depends on the degree of nonnormality of $\ln\left(\hat{\sigma}^2\right)$ and the bias of *se*. The use of prior kurtosis information can only reduce the bias of *se*. Increasing the sample size tends to improve the normality of $\ln\left(\hat{\sigma}^2\right)$ and this highlights the importance of taking a sufficiently large sample from a highly nonnormal distribution.

4. Sample size requirements

Sample size planning is perhaps one of the most important aspects in the design of a study. If the sample size is too small, the width of (4) may be too wide to provide useful information. If the cost of sampling or measuring each sample unit is high, a funding agency may require convincing justification for a proposed sample size. When sampling from animal or human

Table 2
Estimated 95% coverage probabilities of (1) and (4) for nine non-normal distributions

Distribution	n	Eq. (1)	Eq. (4)
Uniform	10	0.993	0.970
	25	0.997	0.950
	50	0.997	0.949
	100	0.997	0.948
Beta(3,3)	10	0.979	0.965
	25	0.981	0.951
	50	0.981	0.949
	100	0.982	0.950
Logistic	10	0.907	0.949
	25	0.892	0.932
	50	0.883	0.932
	100	0.882	0.937
Laplace	10	0.838	0.924
	25	0.814	0.912
	50	0.793	0.916
	100	0.788	0.928
$t(5)$	10	0.874	0.938
	25	0.833	0.912
	50	0.798	0.908
	100	0.784	0.914
Gamma(1,6)	10	0.917	0.955
	25	0.904	0.935
	50	0.896	0.935
	100	0.893	0.939
Beta(1,10)	10	0.829	0.912
	25	0.805	0.912
	50	0.798	0.925
	100	0.793	0.935
Exp	10	0.766	0.888
	25	0.722	0.890
	50	0.697	0.899
	100	0.685	0.917
$\chi^2(1)$	10	0.640	0.850
	25	0.594	0.860
	50	0.565	0.880
	100	0.562	0.900

Table 3
Effect of prior kurtosis information on the 95% coverage probability of (4)

Distribution	γ_4	n_0	$\tilde{\gamma}_4$	n	Coverage
Uniform	1.8	200	1.8	25	0.945
				100	0.948
		500	1.8	25	0.944
				100	0.949
Laplace	6	200	5.6	25	0.958
				100	0.944
		500	5.9	25	0.967
				100	0.951
$t(5)$	9	200	6.1	25	0.971
				100	0.948
		500	7.0	25	0.981
				100	0.962
Gamma(1,6)	4	200	3.8	25	0.960
				100	0.950
		500	3.9	25	0.961
				100	0.952
Exp	9	200	7.9	25	0.959
				100	0.941
		500	8.5	25	0.968
				100	0.952
$\chi^2(1)$	15	200	12.2	25	0.950
				100	0.934
		500	13.6	25	0.964
				100	0.947

populations, the use of an unnecessarily large sample size raises ethical questions if there is any risk of harm or discomfort to the participant.

The relative precision of the interval estimate for σ may be defined as the ratio of the upper to lower endpoints of (4). The following formula closely approximates the sample size needed to obtain a $100(1-\alpha)\%$ confidence interval for σ with desired precision r

$$n \cong \left(\tilde{\gamma}_4 - 1\right)\left\{z_{\alpha/2}/\ln(r)\right\}^2 + 3, \tag{5}$$

where $\tilde{\gamma}_4$ is a planning value of γ_4 obtained from prior research or expert opinion. The reader may verify that setting $\hat{\gamma}_4^* = \tilde{\gamma}_4$ in (4) and using n from (5) will give an upper to lower endpoint ratio that is very close to r. In practice, a range of possible values of $\tilde{\gamma}_4$ might be specified and a conservative choice would be to use the largest value within the specified range.

The inherent difficulty of estimating σ in highly kurtotic distributions is clearly revealed by (5). For instance, to obtain a 95% confidence interval for σ with $r = 1.5$, the required samples size is 22, 50, 190, and 331 for $\tilde{\gamma}_4 = 1.8$, 3, 9, and 15, respectively. Coincidentally, it appears that the sample size needed to obtain an accurate (e.g., $r < 1.5$) estimate of σ is often large enough that (4) will have a true coverage probability that is close to $1 - \alpha$.

5. Example

Kohler (1994, p. 756) describes a quality control study where the fill weight variability of 16-ounce canned peas is assessed at regular intervals. The fill weights for one random sample of $n = 8$ cans is given below.

15.83, 16.01, 16.24, 16.42, 15.33, 15.44, 16.88, 16.31.

From this sample we obtain $\hat{\sigma} = 0.517$ and $\bar{\gamma}_4 = 2.12$. Application of (4) with $\alpha = 0.05$ and no prior kurtosis information gives (0.326, 1.08). The interval is wide because the sample size is small. For subsequent samples, a sample size of about $(2.12 - 1)\{1.96/\ln(1.5)\}^2 + 3 \cong 30$ should give a 95% confidence interval for σ with an upper to lower endpoint ratio that is close to 1.5 if $\bar{\gamma}_4$ is close to 2.12. Because continuous quality control requires repeated sampling over time, it would be wise to maintain a historical record of kurtosis and variance estimates that could be pooled to obtain a prior point estimate of γ_4. This pooled estimate could then be used to improve the performance of (4) in future samples.

6. Concluding remarks

The exact confidence interval for σ given in many text books (1) does not have an asymptotic coverage probability of $1 - \alpha$ in non-mesokurtic ($\gamma_4 \neq 3$) distributions and performs only slightly better than (4) when sampling from a normal distribution. In small samples, tests of normality lack the power to detect the degree of nonnormality that would cause problems with (1). In contrast, tests of normality should have adequate power to detect the type of nonnormality that would cause problems with (4).

In small samples, (4) performs well under moderate departures from normality. As with inferential methods for means, a larger sample size endows (4) with greater protection against nonnormality. If the distribution is highly skewed, a skewness-reducing transformation will decrease the sample size at which the coverage probability of (4) becomes close to $1 - \alpha$. For instance, when sampling from a χ_1^2 distribution, a sample size of about 300 is needed before (4) will have a coverage probability close to $1 - \alpha$. If data from a χ_1^2 distribution are square-root transformed, (4) will then have a coverage probability close to $1 - \alpha$ with a sample size of about 30. Skewness-reducing transformations may also reduce kurtosis which, as can be seen in (5), will reduce the sample size requirement.

We should also consider how (4) compares with a bootstrap confidence interval for σ^2. The percentile bootstrap was found to perform poorly and cannot be recommended. The BC_a method, which is second-order accurate, performed better than the percentile method but worse than (4). For instance, with $n = 25$ and $\alpha = 0.05$, the BC_a method had coverage probabilities of 89.9, 83.4, and 72.8 for the normal, $t(5)$ and $\chi^2(1)$, respectively.

The results of Table 3 clearly illustrate how prior kurtosis information can improve the coverage probability of (4) in nonnormal distributions and should motivate investigators to include kurtosis estimates in their reports for the benefit of future research.

Acknowledgements

The author thanks the Associate Editor and a reviewer for helpful suggestions and E. Seier for assistance with computations.

References

Bartlett, M.S., Kendall, D.G., 1946. The statistical analysis of variance-homogeneity and the logarithmic transformation. J. R. Stat. Soc. Ser. B 8, 128–138.

Cohen, A., 1972. Improved confidence intervals for the variance of a normal distribution. J. Amer. Statist. Assoc. 67, 382–387.

Kohler, H., 1994. Statistical for Business and Economics, third ed. Harper Collins, New York.

Laylard, M.W.J., 1973. Robust large-sample tests for homogeneity of variances. J. Amer. Statist. Assoc. 68, 195–198.

Miller, R.G., 1986. Beyond ANOVA, Basics of Applied Statistics, Wiley, New York.

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

Scheffé, H., 1959. The Analysis of Variance, Wiley, New York.

Shoemaker, L.H., 2003. Fixing the *F* test for equal variances. The Amer. Statistician, 57, 105–114.

Tate, R.F., Klett, G.W., 1959. Optimal confidence intervals for the variance of a normal distribution. J. Amer. Statist. Assoc. 54, 674–682.