Confidence Intervals for the Variance and the Ratio of Variances of Log-Normal Distribution with Known Coefficients of Variation

Sa-aat Niwitpong and Suparat Niwitpong

King Mongkut's University of Technology North Bangkok

Abstract

This paper presents the confidence interval for the variance and the ratio of variances of lognormal distribution with known coefficients of variation. We derived analytic expressions to find the coverage probability and the expected length of the proposed confidence interval.

Mathematics Subject Classification: 62F25

Keywords: Coverage probability; expected length; lognormal distribution

1. Introduction

The lognormal distribution has been widely used for a right skewed data in science, biology and economics. A ratio estimator is much attention in area of bioassay and bioequivalence. Recently, many researchers have been investigated this problem. For example, Lee and Lin (2002) constructed the confidence interval for the normal means by using the generalized confidence interval and the generalized *p*-value proposed by Tsui and Weerahandi (1989). Later, Chen and Zhou (2006) compared several methods for constructing the confidence interval for the ratio of lognormal means. They suggested a modified signed log-likelihood ratio approach which is the best among these confidence intervals. In this paper, we proposed to construct the confidence interval for the lognormal variance and their ratio of variances when the coefficients of variation are known using a simple method. Additionally, we derived analytic expressions to find its coverage probability and its expected length for each interval.

Let $X_i = \left(X_{1i}, X_{2i}, \ldots, X_{n_i}\right), i = 1,2,$, be a random variable having a lognormal distribution, and μ_i and σ_i^2, respectively, are denoted by the mean and the variance of $Y_i = \ln\left(X_i\right) \sim N\left(\mu_i, \sigma_i^2\right)$. The probability density function of X_i, is

$$f\left(x_i,\mu_i,\sigma_i^2\right)=\begin{cases}\dfrac{1}{x_i\sigma_i\sqrt{2\pi}}\exp\left(-\dfrac{\left(\ln\left(x_i\right)-\mu_i\right)^2}{2\sigma_i^2}\right); & if\ \ x_i>0\\ 0 & ;\ \ otherwise.\end{cases}$$

In particular, the mean, variance and the coefficient of variation for lognormal distribution are given by

$$E\left(X_i\right)\ =E\left(\exp\left(Y_i\right)\right)=\exp\left(\mu_i+\frac{\sigma_i^2}{2}\right),Var(X_i)=\exp\left(2\mu_i+\sigma_i^2\right)\left(\exp\left(\sigma_i^2\right)-1\right),$$

$$\tau_i=\sqrt{\exp\left(\sigma_i^2\right)-1},$$

where τ_i denotes the coefficient of variation of X_i which is computed from $\sqrt{Var\left(X_i\right)}/E\left(X_i\right)$. The parameter of interest are

$\delta=k_1\exp\left(2\mu_1+c_1\right)/k_2\exp\left(2\mu_2+c_2\right),\mathrm{k}_i=\exp\left(c_i\right)-1,c_i=\ln(\tau_i^2+1),\mathrm{i}=1,2$ and

$$\theta\ =Var\left(X_1\right)=\exp\left(2\mu_1+\ln(\tau_1^2+1)\right)\left(\exp\left(\ln(\tau_1^2+1)\right)-1\right)$$

$$=k_1\exp\left(2\mu_1+c_1\right),k_1=\exp\left(\ln(\tau_1^2+1)\right)-1$$

when coefficients of variation are known i.e., $\tau_i=\sqrt{\exp\left(\sigma_i^2\right)-1}$ leading to $\sigma_i^2=\ln(\tau_i^2+1)$. Consider $\theta=k_1\exp\left(2\mu_1+c_1\right)$, we take natural logarithm to both sides leading to $\ln(\theta)=\ln(k_1)+\left(2\mu_1+c_1\right)$ and also $\ln(\delta)=\ln(k_1)+\left(2\mu_1+c_1\right)-\ln(k_2)-\left(2\mu_2+c_2\right)=2(\mu_1-\mu_2)+(c_1-c_2)+(\ln(\mathrm{k}_1)-\ln(\mathrm{k}_2))$. We now consider to construct the confidence interval for $\ln(\delta)$ and $\ln(\theta)$ then transform back to the confidence intervals for δ and θ by taking the exponential function to $\ln(\delta)$ and $\ln(\theta)$ respectively.

a) Case 1, when σ_1^2 and σ_2^2 are known

The pivotal quantity for this case is

$$Z=\frac{\ln(\mathrm{k}_1)+(2\bar{Y}_1+c_1)-\ln(\mathrm{k}_2)-(2\bar{Y}_2+c_2)-(\ln(\mathrm{k}_1)+(2\mu_1+c_1)-\ln(\mathrm{k}_2)-(2\mu_2+c_2))}{\sqrt{\dfrac{4\sigma_1^2}{n_1}+\dfrac{4\sigma_2^2}{n_2}}}$$

when $S_i^2=(n_i-1)^{-1}\sum_{i=1}^{n_i}\left(Y_i-\bar{Y}_i\right)^2$ and Z is a standard normal distribution.

$$CI_3=\left[(2\bar{Y}_1+a_1)-(2\bar{Y}_2+a_2)-Z_{1-\alpha/2}\sqrt{\frac{4\sigma_1^2}{n_1}+\frac{4\sigma_2^2}{n_2}},(2\bar{Y}_1+a_1)-(2\bar{Y}_2+a_2)+Z_{1-\alpha/2}\sqrt{\frac{4\sigma_1^2}{n_1}+\frac{4\sigma_2^2}{n_2}}\right]$$

when $a_i = \ln(k_i) + c_i$, i=1,2.

b) Case 2, when σ_1^2 and σ_2^2 are unknown but $\sigma_1^2 = \sigma_2^2$

The pivotal quantity for this case is

$$T_1 = Z = \frac{\ln(k_1) + (2\overline{Y}_1 + c_1) - \ln(k_2) - (2\overline{Y}_2 + c_2) - (\ln(k_1) + (2\mu_1 + c_1) - \ln(k_2) - (2\mu_2 + c_2))}{S_p\sqrt{\frac{4}{n_1} + \frac{4}{n_2}}}$$

when T_1 is the t-distribution with $n_1 + n_2 - 2$ degrees of freedom,

and S_p^2 is the pooled estimate of the sample variance;

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

A 100(1-α) % confidence interval for $\ln(\delta)$ is

$$CI_4 = \left[(2\overline{Y}_1 + a_1) - (2\overline{Y}_2 + a_2) - t_{1-\alpha/2, n1+n2-2} S_p\sqrt{\frac{4}{n_1} + \frac{4}{n_2}}, (\overline{Y}_1 + a_1) - (\overline{Y}_2 + a_2) + t_{1-\alpha/2, n1+n2-2} S_p\sqrt{\frac{4}{n_1} + \frac{4}{n_2}} \right]$$

when $t_{1-\alpha/2}$ is a $(1-\alpha/2)100th$ percentile of the *t*-distribution with n1+n2-2 degrees of freedom and $a_i = \ln(k_i) + c_i$.

c) Case 3, when σ_1^2 and σ_2^2 are unknown but $\sigma_1^2 \neq \sigma_2^2$

The pivotal quantity for this case is

$$T_2 \cong \frac{\ln(k_1) + (2\overline{Y}_1 + c_1) - \ln(k_2) - (2\overline{Y}_2 + c_2) - (\ln(k_1) + (2\mu_1 + c_1) - \ln(k_2) - (2\mu_2 + c_2))}{\sqrt{\frac{4S_1^2}{n1} + \frac{4S_2^2}{n2}}}$$

when T_2 is an approximated t-distribution with

$$v = \frac{(A+B)}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}}, A = \frac{S_1^2}{n_1}, B = \frac{S_2^2}{n_2}$$

degrees of freedom.

A 100(1-α) % confidence interval for $\ln(\delta)$ is

$$CI_5=\begin{bmatrix}\ln(\mathrm{k}_1)+(2\bar{Y}_1+c_1)-\ln(\mathrm{k}_2)-(2\bar{Y}_2+c_2)-t_{1-\alpha/2,\nu}\sqrt{\dfrac{4S_1^2}{n_1}+\dfrac{4S_2^2}{n_2}},\\ \ln(\mathrm{k}_1)+(2\bar{Y}_1+c_1)-\ln(\mathrm{k}_2)-(2\bar{Y}_2+c_2)+t_{1-\alpha/2,\nu}\sqrt{\dfrac{4S_1^2}{n_1}+\dfrac{4S_2^2}{n_2}}\end{bmatrix}.$$

It is easy to see that a 100(1-α) % confidence interval for $\ln(\theta)$ is

$$CI_1=\left[(\ln(\mathrm{k}_1)+c_1)+2\bar{Y}_1-2Z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}}+c_1,\ \ln(\mathrm{k}_1)+c_1+2\bar{Y}_1+2Z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}}+c_1\right]$$

when σ_1^2 is known and

$$CI_2=\left[\ln(\mathrm{k}_1)+c_1+2\bar{Y}_1-2t_{1-\alpha/2,n-1}\sqrt{\frac{S_1^2}{n_1}}+c_1,\ \ln(\mathrm{k}_1)+c_1+2\bar{Y}_1+2t_{1-\alpha/2,n-1}\sqrt{\frac{S_1^2}{n_1}}\right]$$ when σ_1^2 is unknown.

A final process is to use exponential function to transform CI_1, CI_2, CI_3 back to δ, we then have $\exp(CI_1), \exp(CI_2), \exp(CI_3), \exp(CI_4)$ and $\exp(CI_3)$ respectively.

2. Coverage probability and expected length of each confidence interval

In this section, we present the coverage probability and the expected length of each interval.

Theorem 2.1 The coverage probability and the expected length of CI_4 when the variances are equal, $\sigma_1^2=\sigma_2^2$, are respectively

$$E[\Phi(W_1)-\Phi(-W_1)] \quad\text{and}\quad 2^{3/2}d\sigma_1\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{1}{n_1+n_2-2}}\frac{\Gamma(\frac{n_1+n_2-1}{2})}{\Gamma(\frac{n_1+n_2-2}{2})}$$

where $W_1=d_1\sigma_1^{-1}S_p, d_1=t_{1-\alpha/2,n+m-2}$, $\Gamma[.]$ is the gamma function and $\Phi[.]$ is the cumulative distribution function of N (0, 1).
Proof. Similarly to Niwitpong and Niwitpong (2010), from CI_2, we have

$$1-\alpha = P\left[2(\bar{Y}_1-\bar{Y}_2)+(a_1-a_2)-4d_1S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2+(a_1-a_2) < 2(\bar{Y}_1-\bar{Y}_2)+(a_1-a_2)+4d_1S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

$$= P\left[\frac{-2d_1S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}{2\sigma_1\sqrt{n^{-1}+m^{-1}}} < \frac{2(\mu_1-\mu_2)-2(\bar{Y}_1-\bar{Y}_2)}{2\sigma_1\sqrt{n^{-1}+m^{-1}}} < \frac{2d_1S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}{2\sigma_1\sqrt{n^{-1}+m^{-1}}}\right]$$

$$= E[I_{\{-W_1<Z<W_1\}}(\tau)],\ I_{\{-W_1<Z<W_1\}}(\tau) = \begin{cases}1, if\ \tau\in\{-W_1<Z<W_1\}\\ 0, otherwise\end{cases}$$

$$= E[E[I_{\{-W_1<Z<W_1\}}(\tau)] \mid S_p^2 \mid$$

$$= E[\Phi(W_1)-\Phi(-W_1)]$$

where $Z \sim N\,(0;\,1)$.

The expected length of CI_2 is $E\left[4d_1S_p^2\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$,

$$4d_1\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}E[S_p] = 4d_1\sigma_1\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{1}{n_1+n_2-2}}E\left[\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma_1^2}}\right]$$

$$= 4d_1\sigma_1\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{1}{n_1+n_2-2}}E(\sqrt{V})$$

$$= 2^{5/2}d_1\sigma_1\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{1}{n_1+n_2-2}}\frac{\Gamma(\frac{n_1+n_2-1}{2})}{\Gamma(\frac{n_1+n_2-2}{2})}$$

where $V \sim \chi^2_{n_1+n_2-2}$ and $E(\sqrt{V}) = \dfrac{2^{1/2}\Gamma(\frac{1}{2}+\frac{n+m-2}{2})}{\Gamma(\frac{n+m-2}{2})}$, see Casella and Berger (1990). Thus we complete the proof.

Theorem 2.2 The coverage probability and the expected length of CI_5 are respectively

$$E[\Phi(W)-\Phi(-W)]\ and \begin{cases} 2d\sigma_1\sigma_2(n_1n_2)^{-1/2}\delta\sqrt{r_1}F\left[\frac{-1}{2},\frac{n_2-1}{2},\frac{n_2+n_1-2}{2},\frac{r_1-r_2}{r_1}\right], \text{if } r_2<2r_1 \\ 2d\sigma_1\sigma_2(n_1n_2)^{-1/2}\delta\sqrt{r_2}F\left[\frac{-1}{2},\frac{n_1-1}{2},\frac{n_1+n_2-2}{2},\frac{r_2-r_1}{r_2}\right], \text{if } 2r_1\le r_2\end{cases}$$

where

$$W_2 = \frac{d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}{\sqrt{\sigma_1^2 n_1^{-1}+\sigma_2^2 n_2^{-1}}}, d = t_{1-\alpha/2,v}, \delta = \frac{\sqrt{2}\Gamma\left(\frac{n_1+n_2-1}{2}\right)}{\Gamma\left(\frac{n_1+n_2-2}{2}\right)}$$

$$r_1 = \frac{n_2}{\sigma_2^2(n_1-1)}, r_2 = \frac{n_1}{\sigma_1^2(n_2-1)}, \ v = \frac{(A+B)}{\frac{A^2}{n_1-1}+\frac{B^2}{n_2-1}}, A = \frac{S_1^2}{n_1}, B = \frac{S_2^2}{n_2} \ and$$

$E(.)$ is an expectation operator, F $(a; b; c; k)$ is the hypergeometric function,

defined by F $(a; b; c; k)$ = $1+\frac{ab}{c}\frac{k}{1!}+\frac{a(a+1)b(b+1)}{c(c+1)}\frac{k^2}{2!}+...$ where $|k| < 1$, see Press (1966), $\Gamma[.]$ is the gamma function and $\Phi[.]$ is the cumulative distribution function of N (0, 1).

Proof. Since, for normal samples, $\bar{Y}_1, \bar{Y}_2, S_1^2$ and S_2^2 are independent of one another. From CI_3, we have

$$1-\alpha = P\left[2(\bar{Y}_1-\bar{Y}_2)+(a_1-a_2)-2d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}} < 2(\mu_1-\mu_2)+(a_1-a_2) < 2(\bar{Y}_1-\bar{Y}_2)+(a_1-a_2)+2d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

$$= P\left[\frac{-2d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}{2\sqrt{\sigma_1^2 n_1^{-1}+\sigma_2^2 n_2^{-1}}} < \frac{2(\mu_1-\mu_2)-2(\bar{Y}_1-\bar{Y}_2)}{2\sqrt{\sigma_1^2 n_1^{-1}+\sigma_2^2 n_2^{-1}}} < \frac{2d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}{2\sqrt{\sigma_1^2 n_1^{-1}+\sigma_2^2 n_2^{-1}}}\right]$$

$$= P\left[\frac{-d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}{\sqrt{\sigma_1^2 n_1^{-1}+\sigma_2^2 n_2^{-1}}} < Z < \frac{d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}{\sqrt{\sigma_1^2 n_1^{-1}+\sigma_2^2 n_2^{-1}}}\right]$$

$$= E[I_{\{-W_2<Z<W_2\}}(\xi)], I_{\{-W_2<Z<W_2\}}(\xi) = \begin{cases} 1, if\ \xi \in \{-W_2 < Z < W_2\} \\ 0, otherwise \end{cases}$$

$$= E[E[I_{\{-W_2<Z<W_2\}}(\xi)] \mid S], S = (S_1^2, S_2^2)'$$

$$= E[\Phi(W_2) - \Phi(-W_2)]$$

where $Z \sim N$ (0; 1).

The length of $CI_5, L_{CI_5}, is\ 4d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$ and the expected length of L_{CI_5} is

$$= 4d\sigma_1\sigma_2(n_1n_2)^{-1/2}E\left[\sqrt{\frac{n_2S_1^2+n_1S_2^2}{\sigma_1^2\sigma_2^2}}\right]$$

$$4dE[\sqrt{\frac{mS_x^2+nS_y^2}{nm}}] \qquad = 4d\sigma_1\sigma_2(n_1n_2)^{-1/2}E\left[\sqrt{\frac{(\frac{n_2}{n_1-1})}{\sigma_2^2}\frac{(n_1-1)S_1^2}{\sigma_1^2}+\frac{(\frac{n_1}{n_2-1})}{\sigma_1^2}\frac{(n_2-1)S_2^2}{\sigma_2^2}}\right]$$

$$= 4d\sigma_1\sigma_2(n_1n_2)^{-1/2}E[\sqrt{r_1Z_1+r_2Z_2}]$$

$$= \begin{cases} 2d\sigma_1\sigma_2(n_1n_2)^{-1/2}\delta\sqrt{r_1}F\left[\frac{-1}{2},\frac{n_2-1}{2},\frac{n_2+n_1-2}{2},\frac{r_1-r_2}{r_1}\right], \text{if } r_2 < 2r_1 \\ 2d\sigma_1\sigma_2(n_1n_2)^{-1/2}\delta\sqrt{r_2}F\left[\frac{-1}{2},\frac{n_1-1}{2},\frac{n_1+n_2-2}{2},\frac{r_2-r_1}{r_2}\right], \text{if } 2r_1 \le r_2 \end{cases}$$

where $Z_1 = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$, $Z_2 = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$ and for more details of $E[\sqrt{r_1Z_1+r_2Z_2}]$ see Press (1966, pp. 456-458). Thus we complete the proof.

We note here that, it is easy to find the coverage probabilities and the expected lengths of the confidence intervals CI_1, CI_2, CI_3, using the similar method of Theorems 2.1-2.2, so we skip that section.

3. Conclusions

In this paper, we derived the coverage probability and the expected length of CI_4 compared to CI_5. The coverage probabilities of these confidence intervals approach $1-\alpha$, when α is a level of significance and for large sample sizes. The expected lengths for each interval, shown in Theorems 2.1 and 2.2, can be compared. So we do not need to use the simulation to show the results.

Acknowledgements

We are appreciated the funding from Faculty of Applied Sciences (Grants Number 5945108), King Mongkut's University of Technology North Bangkok.

References

Casella, G and Berger, R.L. *Statistical Inference*, California: Duxbury Press, 2002.

Chen, Y. H. and Zhou, X. H. Interval estimates for the ratio and difference of two lognormal means. Statistics in Medicien. **25**(2006): 4099-4113.

Lee, J.C. and Lin, S.H. Generalized confidence intervals for the ratio of means of two normal populations. Journal of Statistical Planning and Inference. **123**(2004): 49-60.

Niwitpong, S. and Niwitpong, S. Confidence interval for the difference of two normal population means with a known ratio of variances. Applied Mathematical Sciences. **4** (2010): 347 – 359.

Press, S.J. A confidence interval comparison of two test procedures for the Behrens-Fisher problems, Journal of the American Statistical Association. **61**(1966), 454-466.

Tsui, K and Weerahandi, S. Generalized p-values in significant testing of hypotheses in the presence of nuisance parameters. The Journal of American Statistical Association. **84**(1989): 602-607.