# Confidence Interval for the Ratio of Lognormal Means When the Coefficients of Variation are Known

Sa-aat Niwitpong
*King Mongkut's University of Technology North Bangkok Thailand*

***Abstract***--This paper presents the confidence interval for the ratio of means of lognormal distribution. We derived analytic expressions to find the coverage probability and the expected length of the proposed confidence interval.

***Keywords***-- Coverage probability, expected length, lognormal distribution

## I. INTRODUCTION

The lognormal distribution has been widely used for a 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 [3] constructed the confidence interval for the normal means by using the generalized confidence interval and the generalized *p*-value proposed by [6]. Later, Chen and Zhou [2] 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 means when the coefficients of variation are known. Additionally, we derived analytic expressions to find its coverage probability and its expected length.

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 $\mu_i$ and $\sigma_i^2$, 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

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

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

where $CV_i$ 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 is $\delta = \exp(\mu_1 + \sigma_1^2/2)/\exp(\mu_2 + \sigma_2^2/2)$, when coefficients of variation are known i.e., $\tau_i = CV_i = \sqrt{\exp\left(\sigma_i^2\right) - 1}$ leading to $\sigma_i^2 = \ln(\tau_i^2 + 1)$ then

$\theta_i = E\left(X_i\right) = \exp\left(\mu_i + \dfrac{\ln(\tau_i^2 + 1)}{2}\right) = \exp\left(\mu_i + c_i\right), \quad c_i = \dfrac{\ln(\tau_i^2 + 1)}{2}$. As a result, the parameter of interest is $\delta = \exp(\mu_1 + c_1)/\exp(\mu_2 + c_2)$. Consider

$\ln(\delta) = \theta_1 - \theta_2, \theta_1 = \mu_1 + \sigma_1^2/2, \theta_2 = \mu_2 + \sigma_2^2/2$ when coefficients of variation are known

$\ln(\delta) = (\mu_1 + c_1) - (\mu_2 + c_2), c_i = \ln(\tau_i^2 + 1)/2.$

We now consider to construct the confidence interval for $\ln(\delta)$ and then transform back to the confidence interval for $\delta$ by taking the exponential function to $\ln(\delta)$.

a) Case 1, when $\sigma_1^2$ and $\sigma_2^2$ are known

The pivotal quantity for this case is

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

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. $CI_1 = \left[ (\bar{Y}_1 + c_1) - (\bar{Y}_2 + c_2) - Z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, (\bar{Y}_1 + c_1) - (\bar{Y}_2 + c_2) + Z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$

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

The pivotal quantity for this case is

$$T_1 = \frac{(\bar{Y}_1 + c_1) - (\bar{Y}_2 + c_2) - ((\mu_1 + c_1) - (\mu_2 + c_2))}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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;

$$\frac{(n_1 - 1)S_1^2 + (\mathrm{n}_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

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

$$CI_2 = \left[ (\bar{Y}_1 + c_1) - (\bar{Y}_2 + c_2) - t_{1-\alpha/2, n1+n2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{Y}_1 + c_1) - (\bar{Y}_2 + c_2) + t_{1-\alpha/2, n1+n2-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

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

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

The pivotal quantity for this case is

$$T_2 = \frac{(\bar{Y}_1 + c_1) - (\bar{Y}_2 + c_2) - ((\mu_1 + c_1) - (\mu_2 + c_2))}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

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

$$CI_3=\left[(\bar{Y}_1+c_1)-(\bar{Y}_2+c_2)-t_{1-\alpha/2,v}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}},(\bar{Y}_1+c_1)-(\bar{Y}_2+c_2)+t_{1-\alpha/2,v}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

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

## II. 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 $CI_2$ when the variances are equal, $\sigma_1^2=\sigma_2^2$, are respectively

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

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 [4], from $CI_2$, we have

$$1-\alpha=P\left[(\bar{Y}_1-\bar{Y}_2)+(c_1-c_2)-d_1S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2+(c_1-c_2)<(\bar{Y}_1-\bar{Y}_2)+(c_1-c_2)+d_1S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

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

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

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

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

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

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

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

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

$$=2^{3/2}d_1\sigma_1\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{1}{n_1+n_2-2}}\frac{\Gamma(\frac{n+m-1}{2})}{\Gamma(\frac{n+m-2}{2})}$$

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

**Theorem 2.2** The coverage probability and the expected length of $CI_3$ 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*

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

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

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

defined by $F$ $(a;\ b;\ c;\ k)$ = $1+\frac{ab}{c}\frac{k}{1!}+\frac{a(a+1)b(b+1)}{c(c+1)}\frac{k^2}{2!}+\ldots$ where $|k| < 1$, see [5], $\Gamma[.]$ 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

$$1-\alpha=P\left[(\bar{Y}_1-\bar{Y}_2)+(c_1-c_2)-d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}<\mu_1-\mu_2+(c_1-c_2)<(\bar{Y}_1-\bar{Y}_2)+(c_1-c_2)+d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right]$$

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

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

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

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

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

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

The length of $CI_2, L_{CI_2}, is\ 2d\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$ and the expected length of $L_{CI_2}$ is

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

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

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

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

where $Z_1 = \frac{(n-1)S_x^2}{\sigma_x^2} \sim \chi_{n-1}^2$, $Z_2 = \frac{(m-1)S_y^2}{\sigma_y^2} \sim \chi_{m-1}^2$ and for more details of $E[\sqrt{r_1Z_1+r_2Z_2}]$ see [5, pp. 456-458]. Thus we complete the proof.

We note here that, it is easy to find the coverage probability and the expected length of the confidence interval $CI_1$, so we skip that section.

## III. CONCLUSIONS

In this paper, we derived the coverage probability and the expected length of $CI_2$ compared to $CI_3$. The coverage probabilities of these confidence intervals approach $1-\alpha$, when $\alpha$ is a level of significance and for large sample sizes. The expected lengths for each interval, shown in Theorems 1 and 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, King Mongkut's University of Technology North Bangkok.

REFERENCES

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

[2] Y.H. Chen and X.H. Zhou, "Interval estimates for the ratio and difference of two lognormal means," Statistics in Medicine, vol 25, pp. 4099-4113, 2006.

[3] J.C. Lee, S.H. Lin, "Generalized confidence intervals for the ratio of means of two normal populations," Journal of Statistical Planning and Inference, vol 123, pp. 49-60, 2004.

[4] S. Niwitpong, S. Niwitpong, "Confidence interval for the difference of two normal population means with a known ratio of variances," Applied Mathematical Sciences, vol 4, pp. 347 – 359, 2010

[5] S.J. Press, "A confidence interval comparison of two test procedures for the Behrens-Fisher problems," Journal of the American Statistical Association, vol 61, pp. 454-66, 1966.

[6] K. Tsui and S. Weerahandi, "Generalized p-values in significant testing of hypotheses in the presence of nuisance parameters," The Journal of American Statistical Association," vol 84, pp. 602-607, 1989..