



*Original Article*

## The variable for the generalized confidence interval for the lognormal mean

Thongkam Maiklad \*

*Department of Applied Statistics, Faculty of Applied Science,  
King Mongkut's University of Technology North Bangkok, Bang Sue, Bangkok 10800, Thailand.*

Received 18 September 2007; Accepted 28 July 2008

### Abstract

The purpose of this paper is to propose a new variable for the generalized confidence interval method to estimate the confidence interval of the lognormal mean. In order to evaluate the efficiency of this new method, here called t-generalized method, a simulation study was conducted to examine and compare the coverage probability, interval width, and relative bias of this new method and three other methods, the generalized confidence method of Krishnamoorthy and Mathew, the Modified Cox method, and the Angus's conservative method. The results show that at small sample sizes with large variances, only the t-generalized method and generalized confidence method of Krishnamoorthy and Mathew provide coverage probabilities greater than the nominal level. The t-generalized method is more accurate with a shorter confidence interval than the old generalized confidence method in the case of small sample sizes with large variances.

**Keywords:** coverage probability, interval width, relative bias, generalized confidence interval, modified cox, Angus's conservative, t-generalized

### 1. Introduction

The lognormal distribution is a skewed distribution which is widely used for analyzing the data sets where most of the observations are small, but with a few very large values. Such data, for example, may be the costs of a hospital stay, the incomes of individuals, the height of flood in a river, the amount of Hartmonelly hyaline per gram of soil.

Let  $X$  be a random variable having a lognormal distribution,  $\sim \text{lognormal}(\mu, \sigma^2)$ . Then  $Y = \log(X)$  has a normal distribution,  $N(\mu, \sigma^2)$ . The density of  $X$  is

$$f(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-(\ln x - \mu)^2 / 2\sigma^2),$$
$$0 < x < \infty, 0 < \sigma < \infty.$$

The mean and variance of  $X$  are  $E(X) = \theta = \exp(\mu + \sigma^2 / 2)$  and  $V(X) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)$ , respectively. The

most commonly used method for obtaining the confidence limits for  $\theta$  is the so-called naïve transformation method. This method constructs a confidence interval for  $\exp(\mu)$  which is the median of  $X$ . The result of this method is tolerably accurate for  $\theta$  when  $\sigma^2$  is relatively small, but becomes intolerably inaccurate as  $\sigma^2$  increases. The accuracy gets worse as the sample size increases. In 1957, Aitchison and Brown (Land, 1972) suggested an approximate confidence interval method or transformation method which should converge to the exact limits only when the sample size becomes infinitely large. Zhou and Gao (1997) compared the coverage probabilities of from the naïve transformation method, the Cox method, the Angus's conservative method and the parametric bootstrap. The simulation results showed that the parametric bootstrap method was the most appropriate method for small variances, whereas Angus's conservative method always gave coverage probabilities more than the nominal level. After Weerahandi (1993) developed the generalized confidence interval, Krishnamoorthy and Mathew (2003) compared the upper limit of the 95% confidence interval of  $\ln \theta$  from this method with the Land formula and the parametric bootstrap method. The result showed

\*Corresponding author.

Email address: [tkm@kmutnb.ac.th](mailto:tkm@kmutnb.ac.th)

that the generalized confidence limit and the confidence limit obtained by Land's formula practically coincide. The coverage probability from the generalized confidence interval method is always close to the nominal level. When sample sizes were small, the parametric bootstrap gave coverage probabilities less than the nominal level. Olsson (2005) suggested the modified Cox method and used simulation studies to compare the coverage probabilities of this new method with the naïve transformation, the Cox, the generalized confidence interval and the large sample methods. The results showed that only the coverage probabilities from the Modified Cox method and the generalized confidence interval method are close to the nominal level, but the interval widths from the Modified Cox method are larger.

The purpose of this paper is to suggest another method of constructing the confidence interval of the lognormal mean. This new method, t-generalized method, is derived from Weerahandi's generalized confidence interval. In Section 4, the efficiency of estimation is evaluated by a simulation study. The coverage probability, interval width and relative bias from the t-generalized method is compared to the generalized confidence interval method of Krishnamoorthy and Mathew, the Angus's conservative method and the Modified Cox method.

**2. The four methods of confidence interval estimation**

In this section, the four methods for constructing two-sided  $(1-\alpha)$  100% confidence intervals for  $\theta = \exp(\mu + \sigma^2/2)$ , the lognormal mean, is reviewed.

**2.1 The generalized confidence interval method of Krishnamoorthy and Mathew**

Krishnamoorthy and Mathew (2003) recommended an exact confidence interval using the ideas of the generalized confidence interval of Weerahandi (1993). Their confidence interval for the lognormal mean has the following algorithm:

For a given lognormal data set  $x_1, \dots, x_n$ , let  $y_i = \ln(x_i)$ ;  $i = 1, \dots, n$ .

Compute  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  and  $s^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$ .

Generate  $Z \sim N(0, 1)$  and  $U^2 \sim \chi^2_{(n-1)}$ .

Set  $R = \bar{y} \cdot \frac{Z}{U/\sqrt{n-1}} \frac{s}{\sqrt{n}} + \frac{s^2}{2U^2/(n-1)}$ .

The  $100(\alpha/2)^{th}$  and the  $(1-\alpha/2)^{th}$  percentile of R are the lower and the upper limit, respectively, of the  $(1-\alpha)$  100% confidence interval for  $\ln \theta$ .

**2.2 Angus's conservative method**

The Angus's conservative method for construction of a confidence interval for  $\ln \theta$  is based on the following approximate pivotal statistic:

$$V(\theta) = \frac{\sqrt{n}(\bar{Y} + S^2/2 - \ln \theta)}{\sqrt{S^2(1 + S^2/2)}} \tag{2.2.1}$$

where  $\bar{Y}$  and  $S^2$  are the mean and variance of  $Y = \log(X)$ , respectively.

In a finite sample, the approximate pivotal statistic in Equation (2.2.1) has the same distribution as

$$T(\sigma) = \frac{N + \sigma \frac{\sqrt{n}}{2} \left( \frac{\chi^2_{(n-1)}}{n-1} - 1 \right)}{\sqrt{\frac{\chi^2_{(n-1)}}{n-1} \left( 1 + \frac{\sigma^2}{2} \frac{\chi^2_{(n-1)}}{n-1} \right)}}, \tag{2.2.2}$$

where  $N$  and  $\chi^2_{(n-1)}$  are independent,  $N$  has a the standard normal distribution, and  $\chi^2_{(n-1)}$  is a  $\chi^2$ -distribution with  $n-1$  degrees of freedom. Let  $F(z; \sigma)$  be the cumulative distribution of  $T(\sigma)$ . For fixed  $z$  and fixed  $n > 2$ ,  $F(z; \sigma)$  is monotone increasing in  $\sigma$ . We have

$$\lim_{\sigma \rightarrow 0} T(\sigma) = \frac{N}{\sqrt{\chi^2_{(n-1)} / (n-1)}} \sim t_{(n-1)}$$

and  $\lim_{\sigma \rightarrow \infty} T(\sigma) = \sqrt{\frac{n}{2} \left( 1 - \frac{n-1}{\chi^2_{(n-1)}} \right)}$

hence

$$\inf_{\sigma > 0} F(z; \sigma) = \Pr(t_{(n-1)} \leq z) \text{ and } \sup_{\sigma > 0} F(z; \sigma) =$$

$$\Pr \left( \sqrt{\frac{n}{2} \left( 1 - \frac{n-1}{\chi^2_{(n-1)}} \right)} \leq z \right).$$

Let  $t_{1-\alpha, n-1}$  be the  $(1-\alpha)^{th}$  quantile of a  $t$  distribution with

$n-1$  degrees of freedom, and let  $q_{\alpha}(n-1) = \sqrt{\frac{n}{2} \left( \frac{n-1}{\chi^2_{\alpha, n-1}} - 1 \right)}$ ,

where  $\chi^2_{\alpha, n-1}$  denote the  $\alpha^{th}$  quantile of a  $\chi^2$ -distribution with  $n-1$  degrees of freedom.

Then a conservative  $(1-\alpha)$  100% confidence interval for  $\ln \theta$  is

**2.3 The Modified Cox method**

Olsson (2005) modified the confidence interval of the Cox method by using the *t* - distribution instead of the standard normal distribution, so that the confidence interval for (1- $\alpha$ ) 100% confidence interval for  $\ln \theta$  is

$$\left( \bar{y} - \frac{s^2}{2} - t_{1-\alpha/2, n-1} \sqrt{\frac{s^2}{n} + \frac{s^4}{2(n-1)}}, \bar{y} - \frac{s^2}{2} + t_{1-\alpha/2, n-1} \sqrt{\frac{s^2}{n} + \frac{s^4}{2(n-1)}} \right),$$

where  $t_{1-\alpha/2, n-1}$  is the  $(1-\alpha/2)^{th}$  quantile of a *t* distribution with *n*-1 degrees of freedom.

**2.4 The proposed method or t-generalized method**

Let  $X_1, \dots, X_n$  be a sample from lognormal( $\mu, \sigma^2$ ), and let  $Y_i = \log(X_i), i = 1, \dots, n$ . We shall develop the generalized pivotal quantity for  $\ln \xi\theta$  based on the sufficient

statistics  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$  and  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ . We denote  $\bar{y}$  and  $S^2$ , the observed values of  $\bar{Y}$  and  $S^2$ , respectively. In order to estimate the mean of the lognormal distribution, we shall first define a generalized pivotal quantity  $T_j$  that is a function of the random variables  $\bar{Y}$  and  $S^2$ , and their observed values  $\bar{y}$  and  $S^2$ , where  $T_j$  has following two properties referring to Weerahandi(1995):

- (1)  $T_j$  has a probability distribution that is free of unknown parameters.
- (2) The observed value of  $T_j$  is free of any unknown parameters.

We shall now construct a generalized pivotal statistic satisfying the condition above. Consider the identity

$$\mu + \frac{\sigma^2}{2} \equiv \bar{y} - \frac{\bar{Y} - \mu}{S/\sqrt{n}} \times \frac{s}{\sqrt{n}} + \frac{\sigma^2(n-1)s^2}{2(n-1)S^2}$$

Table 1 Coverage probabilities from four methods of calculating 2-sided 95% confidence intervals.

n	Method	s <sup>2</sup> = 0.1	s <sup>2</sup> = 0.5	s <sup>2</sup> = 1	s <sup>2</sup> = 2	s <sup>2</sup> = 5	s <sup>2</sup> = 10	s <sup>2</sup> = 15
10	Modified Cox	0.9760	0.9660	0.9578	0.9444	0.9216	0.9108	0.9052
	Angus's Conservative	0.9920	0.9880	0.9836	0.9712	0.9554	0.9404	0.9350
	Generalized Confidence Interval	0.9748	0.9686	0.9638	0.9594	0.9570	0.9540	0.9530
	t- Generalized Confidence	0.9796	0.9810	0.9778	0.9754	0.9706	0.9660	0.9634
15	Modified Cox	0.9746	0.9678	0.9602	0.9482	0.9354	0.9278	0.9246
	Angus's Conservative	0.9936	0.9912	0.9870	0.9798	0.9690	0.9624	0.9612
	Generalized Confidence Interval	0.9742	0.9672	0.9660	0.9608	0.9534	0.9540	0.9544
	t- Generalized Confidence	0.9734	0.9756	0.9740	0.9708	0.9652	0.9630	0.9602
20	Modified Cox	0.9678	0.9624	0.9576	0.9474	0.9360	0.9282	0.9246
	Angus's Conservative	0.9896	0.9920	0.9880	0.9828	0.9770	0.9710	0.9684
	Generalized Confidence Interval	0.9714	0.9666	0.9630	0.9598	0.9580	0.9566	0.9558
	The Adjusted Generalized Confidence	0.9726	0.9702	0.9694	0.9672	0.9628	0.9608	0.9572
30	Modified Cox	0.9730	0.9676	0.9618	0.9532	0.9416	0.9366	0.9322
	Angus's Conservative	0.9914	0.9934	0.9940	0.9916	0.9876	0.9832	0.9824
	Generalized Confidence Interval	0.9720	0.9710	0.9646	0.9632	0.9610	0.9572	0.9572
	t- Generalized Confidence	0.9748	0.9708	0.9688	0.9648	0.9626	0.9602	0.9586
50	Modified Cox	0.9742	0.9692	0.9648	0.9596	0.9516	0.9482	0.9452
	Angus's Conservative	0.9906	0.9928	0.9932	0.9930	0.9908	0.9900	0.9890
	Generalized Confidence Interval	0.9724	0.9720	0.9668	0.9646	0.9592	0.9550	0.9550
	t- Generalized Confidence	0.9744	0.9728	0.9706	0.9662	0.9620	0.9584	0.9562
100	Modified Cox	0.9696	0.9666	0.9618	0.9576	0.9514	0.9506	0.9496
	Angus's Conservative	0.9870	0.9894	0.9904	0.9918	0.9932	0.9920	0.9914
	Generalized Confidence Interval	0.9698	0.9664	0.9618	0.9566	0.9532	0.6544	0.9540
	t- Generalized Confidence	0.9698	0.9658	0.9640	0.9590	0.9544	0.9560	0.9562
200	Modified Cox	0.9748	0.9704	0.9684	0.9624	0.9568	0.9512	0.9490
	Angus's Conservative	0.9888	0.9898	0.9906	0.9908	0.9910	0.9904	0.9894
	Generalized Confidence Interval	0.9730	0.9700	0.9656	0.9610	0.9580	0.9552	0.9546
	t- Generalized Confidence	0.9738	0.9714	0.9672	0.9638	0.9590	0.9568	0.9536



are respectively the lower and the upper limit, respectively, of the  $(1-\alpha)$  100% confidence interval for  $\ln \theta$ .

**3. Simulation framework**

A simulation study was conducted to compare the coverage probability, interval width, and relative bias of the four methods for constructing two-sided 95% confidence intervals for  $\ln \theta$ . Sample sizes of  $n = 10, 15, 20, 30, 50, 100,$  and  $200$  were used while values of the variance,  $\sigma^2$ , used were  $0.1, 0.5, 1, 2, 5, 10$  and  $15$ . To avoid losing generality,  $\mu$ , the mean of  $Y$ , are set  $\mu = -\sigma^2/2$ , so that  $\ln \theta = 0$ . For each parameter configuration,  $5,000$  random samples from the lognormal distribution were generated. The loop of  $m$  in the generalized confidence interval method of Krishna-moorthy and Mathew and the t-generalized method was  $10,000$ .

The criteria for comparison are coverage probability, average length of the intervals, and how the intervals fail to cover the true parameter  $\theta$  defined as relative bias, which is

$$\text{Relative bias} = \frac{(\%CI < \theta) - (\%CI > \theta)}{(\%CI < \theta) + (\%CI > \theta)}$$

where  $\%CI < \theta$  is the percentage of the intervals falling below the true parameter  $\theta$  and  $\%CI > \theta$  is the percentage of the intervals falling above the true parameter.

**4. Simulation results**

The coverage probabilities as shown in Table 1 from the t-generalized method are significantly higher than those from the generalized confidence interval method of Krishna-moorthy and Mathew. However, the t-generalized method provides average lengths values that are shorter than those from the generalized confidence interval method of Krishna-moorthy and Mathew (Table 2). When the sample size is small with a very large variance, the Modified Cox and Angus's conservative methods yield coverage probabilities significantly lower than  $0.95$ . Even though for sample sizes not less

Table 3. Relative bias obtained from four methods of calculating 2-sided 95% confidence intervals

n	Method	s <sup>2</sup> = 0.1	s <sup>2</sup> = 0.5	s <sup>2</sup> = 1	s <sup>2</sup> = 2	s <sup>2</sup> = 5	s <sup>2</sup> = 10	s <sup>2</sup> = 15
10	Modified Cox	-0.62	-0.92	-0.99	-1	-1	-1	-1
	Angus's Conservative	0.15	-0.73	-0.98	-1	-1	-1	-1
	Generalized Confidence Interval	0.17	0.36	0.35	0.32	0.2	0.16	0.123
	t- Generalized Confidence	-0.29	-0.49	-0.46	-0.48	-0.32	-0.21	-0.126
15	Modified Cox	-0.67	-0.91	-0.98	-1	-1	-1	-1
	Angus's Conservative	0.31	-0.68	-0.94	-0.98	-1	-1	-1
	Generalized Confidence Interval	-0.12	0.10	0.15	0.16	0.15	0.13	0.13
	t- Generalized Confidence	-0.44	-0.65	-0.54	-0.45	-0.28	-0.13	-0.03
20	Modified Cox	-0.49	-0.84	-0.94	-0.98	-1	-1	-1
	Angus's Conservative	0.58	-0.20	-0.80	-0.93	-1	-1	-1
	Generalized Confidence Interval	-0.06	-0.02	-0.01	0.02	0.09	0.06	0.04
	t- Generalized Confidence	-0.34	-0.41	-0.52	-0.43	-0.21	-0.13	-0.08
30	Modified Cox	-0.42	-0.72	-0.86	-0.91	-0.96	-0.98	-0.99
	Angus's Conservative	0.86	0.45	0.07	-0.29	-0.71	-0.86	-0.93
	Generalized Confidence Interval	0.04	0.10	0.15	0.11	0.05	0.07	0.09
	t- Generalized Confidence	-0.24	-0.27	-0.28	-0.26	-0.21	-0.05	0
50	Modified Cox	-0.27	-0.56	-0.67	-0.76	-0.80	-0.83	-0.85
	Angus's Conservative	1	0.89	0.76	0.49	0.22	-0.04	-0.20
	Generalized Confidence Interval	0.03	0.08	0.12	0.20	0.20	0.19	0.18
	The Adjusted Generalized Confidence	-0.15	-0.29	-0.22	-0.14	-0.02	0.04	0.08
100	Modified Cox	-0.14	-0.39	-0.51	-0.67	-0.78	-0.79	-0.80
	Angus's Conservative	1	1	0.96	0.76	0.71	0.30	0.26
	Generalized Confidence Interval	0.09	0.06	0.03	0.01	-0.02	-0.05	-0.05
	t- Generalized Confidence	-0.05	-0.12	-0.20	-0.20	-0.16	-0.13	-0.10
200	Modified Cox	-0.11	-0.31	-0.40	-0.52	-0.59	-0.61	0.59
	Angus's Conservative	1	1	1	1	1	1	1
	Generalized Confidence Interval	0.07	0.05	0.06	0.03	-0.02	-0.06	-0.09
	t- Generalized Confidence	-0.04	-0.11	-0.14	-0.18	-0.11	-0.12	-0.08

than 15 the Angus's conservative method gave the greatest coverage probabilities and, it provided the largest average length among the four methods. The results show that the relative bias of the Modified Cox method has negative values in every situation (table 3). If the sample size is small and the variance is large, the Modified Cox method and Angus's conservative method have a relative bias value of -1. At  $n = 200$  the Angus's conservative method provided +1 for the value of the relative bias. The relative bias of the t-generalized method and the generalized confidence interval method of Krishnamoorthy and Mathew are not far from 0. Almost all of the relative biases of the t-generalized method are less than the relative biases of the generalized confidence interval method of Krishnamoorthy and Mathew.

### 5. Concluding remarks

The results for the generalized confidence interval method of Krishnamoorthy and Mathew always provide coverage probabilities more than the nominal level, whereas Olsson (2005) and Krishnamoorthy and Mathew (2003) found that this method gave coverage probabilities close to the nominal level. The coverage probabilities from the Angus's method are significantly smaller than the nominal level, whenever the sample size is 10 and . This observation indicates, contrary to the studies of Zhou and Gao (1997), that the coverage probabilities based on the Angus's method are always over the nominal level. To estimate the confidence interval of the lognormal mean based on the Weerahandi's method, the suggested pivotal statistic should give coverage

probabilities not less than the nominal level with smaller interval widths.

### References

- Angus, J.E. 1998. Inferences on the lognormal mean for complete samples. *Communications in Statistics-Simulation and Computation*, 17, 1307-1331.
- Krishnamoorthy, K. and Mathew, T. 2003. Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. *Journal of Statistical Planning and Inference*, 115, 103-121.
- Land, C. 1972. An Evaluation of Approximate Confidence Interval Estimation Methods for Lognormal Means. *Technometrics*, 14, 145-158.
- Olsson, U. 2005. Confidence Intervals for the Mean of a Lognormal Distribution. *Journal of Statistics Education*, Online at: <http://www.amstat.org/publications/jse/>, 13, Number 1.
- Patterson, R.L. 1966 Difficulties involved in the estimation of a population mean using transformed sample data. *Technometrics*, 8, 535-537.
- Weerahandi, S. 1993 Generalized Confidence Intervals. *Journal of the American Statistical Association*, 88, 899-905.
- Weerahandi, S. 1995 *Exact Statistical Methods for Data Analysis*. Springer-Verlag, New York, U.S.A.
- Zhou, X-H. and Gao, S. 1997 Confidence Intervals for the Lognormal Mean. *Statistics in Medicine*, 16, 783-790.