Thailand Statistician
October 2020; 18(4): 439-456
http://statassoc.or.th
Contributed paper

Statistical Estimation of Mean of Delta-Lognormal Distribution

Patcharee Maneerat [a][b], Sa-Aat Niwitpong*[b] and Suparat Niwitpong [b]
[a] Department of Mathematics, Uttaradit Rajabhat University, Uttaradit, Thailand.
[b] Department of Applied Statistics, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand.
*Corresponding author; e-mail: sa-aat.n@sci.kmutnb.ac.th

Received: 23 May 2019
Revised: 30 August 2019
Accepted: 7 September 2019

Abstract

In this article, we propose the following approaches derive from the method of variance estimate recovery based on the variance-stabilizing transformation (MOVER-VST), Wilson score (MOVER-Wilson) and Jeffreys (MOVER-Jeffreys) compared with the generalized confidence interval (GCI) to develop the statistical estimation being confidence intervals for single and difference between two means in delta-lognormal distribution. Monte Carlo simulation is used as a technique to evaluate the performance of these confidence intervals in terms of coverage probability and average length. In simulation study, the numerical results of single mean showed that the MOVER-VST and MOVER-Wilson can be considered as the recommended CIs to estimate the delta-lognormal mean in the important cases. For the difference between two delta-lognormal means, numerical computation indicated that the MOVER-Jeffreys achieved the given target when the dispersion was not large for the large probability of additional zero. For application, we illustrate the presented confidence intervals with real world data sets in several fields: the airborne chlorine record for environmental problem, the red cod density for fishery survey and the distance traveled of mice for biology.

Keywords: Generalized confidence interval, Jeffreys interval, method of variance estimates recovery, variance-stabilizing transformation, Wilson score interval.

1. Introduction

Aitchison (1955) first introduced a delta-lognormal distribution. It is represented as the data containing both zero and positive values. The number of zero observations have a binomial distribution with the probability δ where $0<\delta<1$. The non-zero values are derived from a lognormal distribution with the remaining probability $1-\delta$. Many fields utilize this distribution to apply in several real-word situations such as medicine, environment and fishery. For example, it was used to study the medical cost in patient groups where zero corresponded to cases of no health care cost (Zhou and Tu 2000; Tierney et al. 2003; Tian 2005), to analyze airborne chlorine quantity for indicating air contaminant at industrial sites, United State where zeros corresponded to cases of the evaluation cannot reveal the concentration in lunch room and break area (Owen and DeRouen 1980; Tian 2005; Tian and Wu 2006), the fatty acids of captive seabirds where zeros corresponded to cases of absent diet (Stewart and Field 2011), and to estimate the red cod density for assessing their abundance where zeros corresponded to cases of trawls are blank (Pennington 1983; Smith 1988; Smith 1990; Lo et al. 1992; Fletcher 2008; Wu and Hsieh 2014).

A mean is considered as one of the parameter of interest representing the population. For two populations, the difference and ratio between two means are also the parameter of interest to compare two quantity in the same distribution. These parameters regard a statistic gathering the population information. Moreover, the mean has been utilized in many applications, such as environment, pharmacokinetics, medicine and fishery. For example, it is used to estimate the air lead levels at the Alma American Labs, Colorado (Zou et al. 2009b), to assess the maximum concentration from alcohol interaction study in men (Tian and Wu 2007; Krishnamoorthy and Oral 2017), to analyze the medical charge data from the Regenstrief Medical Record System (Tian and Wu 2007; Zhou et al. 1997) and to evaluate the fish densities from a fisheries New Zealand trawl survey (Fletcher 2008; Wu and Hsieh 2014).

These situations confirm that the delta-lognormal distribution is widely applied in several fields. Importantly, one of the statistical inferences is interval estimation providing the information with respect to the population more than point estimate (Casella and Berger 2002). For these reason, researchers have studied interval estimates for the parameters, including mean, variance and coefficient of variation of delta-lognormal population in the different methods. For instance, Zhou and Tu (2000) proposed a percentile-t bootstrap interval based on sufficient statistics and two likelihood-based confidence intervals (CIs) for the mean of diagnostic test charges data containing zeros, they found that the bootstrap performed well for small skewness and sample size especially. Tian (2005) presented the concepts of generalized test variables and generalized pivotal quantities (GPQs) on the zero-inflated lognormal mean, whereas his GCI cannot satisfy a desired value when sample size was large. Tian and Wu (2006) considered the adjusted signed log-likelihood ratio statistic to produce for the mean of lognormal data with excess zeros, although it only provided a good coverage probability (CP) performance. Chen and Zhou (2006) presented CIs for the ratio or difference of means in lognormal populations with zeros using four methods: a true generalized pivotal (GP), an approximate GP, a signed log-likelihood ratio (SLLR) and modified SLLR method. Fletcher (2008) proposed a profile-likelihood for delta-lognormal mean. Harvey and Merwe (2012) presented a Bayesian approach with both the equal-tail and highest posterior density (HPD) using the different priors for single mean, variance and the ratio of delta-lognormal means, however they have studied in the situation of having large zero observations. Li et al. (2013) found that the fiducial approach had high performance in terms of coverage probability and low bias in CIs for mean of lognormal data with excess zeros. Later, Wu and Hsieh (2014) adjusted asymptotic GPQ to construct GCI based on variance stabilized transformation (VST) for the mean of the delta-lognormal population, their research results reveals that the GCI lengths were not less than the existing profile likelihood of Fletcher (2008). Rosales and Naranjo (2016) studied the pooled-t, Weleh's t and Wilson-Hogdes-Lehmann in the constructing of CIs for mean difference between two delta-lognormal data. Recently, Hasan and Krishnamoorthy (2018) improved the fiducial CIs for the mean and a percentile based on delta-lognormal distribution, although their numerical evaluation have not reported for occurring zero values with the probability of 0.5.

As mentioned above, many studies have examined focusing on CIs for delta-lognormal mean, meanwhile they extended to study the mean difference. Unfortunately, these research studies still had restriction in a few aspects. To our knowledge, this paper considers interval estimates for the single and difference between two means of delta-lognormal distributions using three proposed CIs: the method of variance estimates recovery based on the variance-stabilizing transformation (MOVER-

VST), Wilson (MOVER-Wilson) and Jeffreys (MOVER-Jeffreys). These proposed CIs are compared with the existing GCI of Wu and Hsieh (2014). The theories and methods are elaborated to establish CIs for delta-lognormal mean in Section 2. Those methodologies are expanded to construct CIs for the difference between two means in Section 3. The CI performances are assessed by the part of simulation studies in Section 4. Real-world data are used to illustrate with our presented methods in Section 5. This paper is closed with the discussion and conclusions in Section 6.

2. Methods for Constructing CIs for Delta-Lognormal Mean

Let a random sample $X = (X_1, X_2, ..., X_n)$ be drawn from delta-lognormal distribution with the parameters μ, σ^2 and δ, denoted as $\Delta(\mu, \sigma^2, \delta)$. Aitchison (1955) presented the distribution function of X, given by

$$G(x;\mu,\sigma^2,\delta) = \begin{cases} \delta & ; x = 0 \\ \delta + (1-\delta)F(x;\mu,\sigma^2) & ; x > 0, \end{cases}$$

where the number of zero observations has the binomial distribution, denoted to be $n_{(0)} \sim B(n,\delta)$, $\delta = P(X = 0)$. $F(x;\mu,\sigma^2)$ stands for the cumulative distribution function of lognormal distribution and $Y = \ln X \sim N(\mu,\sigma^2)$; μ and σ^2 are the mean and variance of Y, respectively. Then, the probability density function (pdf) of X is

$$f(x;\mu,\sigma^2) = \begin{cases} \dfrac{1}{x\sqrt{2\pi\sigma^2}} \exp(-\dfrac{1}{2\sigma^2}(\ln x - \mu)^2) & ; x > 0 \\ 0 & ; \text{otherwise}. \end{cases}$$

The maximum likelihood estimates (MLEs) of μ, σ^2 and δ are proved as $\hat{\mu} = \frac{1}{n_{(1)}}\sum_{i=1}^{n_{(1)}} y_i$, $\hat{\sigma}^2 = \frac{1}{n_{(1)}}\sum_{i=1}^{n_{(1)}}(y_i - \hat{\mu})^2$ and $\hat{\delta} = \frac{n_{(0)}}{n}$; $n_{(0)} + n_{(1)} = n$ where $n_{(1)}$ are the number of positive observed values. The population mean, variance and coefficient of variation of X are

$$\vartheta = (1-\delta)\exp\left(\mu + \frac{\sigma^2}{2}\right),$$

$$\kappa = (1-\delta)\exp(2\mu + \sigma^2)\left[\exp(\sigma^2) + \delta - 1\right],$$

$$\phi = \sqrt{\frac{\exp(\sigma^2) + \delta - 1}{1-\delta}}.$$

In this study, CIs for ϑ are investigated using the following methods.

2.1. GCI for the mean

The GCI first presented by Weerahandi (1993) being the general approach to establish CI based on the GPQ concept defined as follows. Let $X = (X_1, X_2, ..., X_n)$ be a random variables with the probability density function $f_X(x; \vartheta, \delta)$. The ϑ and σ^2 are the parameter of interest and the vector of nuisance parameter, respectively. Given $x = (x_1, x_2, ..., x_n)$ be a observed values of X. The GCI for ϑ is calculated by using the percentile of GPQ $R(X; x, \vartheta, \delta)$ as if the following conditions are satisfied as

(i) Given X, the distribution of $R(X; x, \vartheta, \delta)$ is free of all unknown parameters.

(ii) The observed value of $R(X; x, \vartheta, \delta)$, depend on the parameter of interest ϑ.

If $R(X; x, \vartheta, \delta)$ is followed by the conditions (i) and (ii), the $100(1-\zeta)\%$ GCI for ϑ is $[R_{\zeta/2}, R_{1-\zeta/2}]$ where R_ζ be the ζ^{th} percentile of $R(X; x, \vartheta, \delta)$. For μ and σ^2, Krishnamoorthy and Mathew (2003) proposed the GPQs for both, given by

$$R_\mu = \hat{\mu} - \frac{W}{\sqrt{n_{(1)}}}\sqrt{\frac{(n_{(1)}-1)\hat{\sigma}^2}{U}},$$

$$R_{\sigma^2} = \frac{(n_{(1)}-1)\hat{\sigma}^2}{U},$$

where $W = (\hat{\mu}-\mu)\Big/\sqrt{\frac{(n_{(1)}-1)\hat{\sigma}^2}{n_{(1)}U}} \sim N(0,1)$ and $U = \frac{(n_{(1)}-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n_{(1)}-1}$. According to Wu and Hsieh (2014), the GPQ for δ is

$$R_\delta = \sin^2\left[\arcsin\sqrt{\hat{\delta}} - \frac{Z}{2\sqrt{n}}\right],$$

where $Z = 2\sqrt{n}\left(\arcsin\sqrt{\hat{\delta}} - \arcsin\sqrt{\delta}\right) \xrightarrow{d} N(0,1)$. By the information of three pivots, the equivalent of ϑ is defined as

$$R_\vartheta = (1-R_\delta)\exp\left(R_\mu + \frac{R_{\sigma^2}}{2}\right).$$

It can be expressed as

$$R_\vartheta = \left\{1-\sin^2\left[\arcsin\sqrt{\hat{\delta}} - \frac{Z}{2\sqrt{n}}\right]\right\}\exp\left[\hat{\mu} - \frac{W}{\sqrt{n_{(1)}}}\sqrt{\frac{(n_{(1)}-1)\hat{\sigma}^2}{U}} + \frac{(n_{(1)}-1)\hat{\sigma}^2}{2U}\right], \quad (1)$$

where W, U and Z are independent random variables. The GPQ in (1) satisfies the two conditions of Weerahandi (1993) for being a GPQ. Then, the $100(1-\zeta)\%$ GCI for ϑ is given as

$$CI_{gci} = [L_{gci}, U_{gci}] = [R_\vartheta(\zeta/2), R_\vartheta(1-\zeta/2)],$$

where $R_\vartheta(\zeta)$ denotes the ζ^{th} percentile of R_ϑ.

2.2. MOVER for the mean

According to Donner and Zou (2012), let $\hat{\theta}_k$ be a estimate of θ_k, $k=1,2$. The θ_1 and θ_2 are independent, so the lower limit for $\theta_1+\theta_2$ using central limit theorem (CLT) can be written as

$$L_{\theta_1+\theta_2} = (\hat{\theta}_1+\hat{\theta}_2) + V_{\zeta/2}\sqrt{var(\hat{\theta}_1)+var(\hat{\theta}_2)},$$

where V_ζ be the ζ^{th} percentile of standard normal distribution. Let (l_k, u_k) be the CIs for θ_k. Notice that $L_{\theta_1+\theta_2}$ can be closed to l_1+l_2 more than $\hat{\theta}_1+\hat{\theta}_2$. By CLT, the estimate variance at $\theta_k = l_k$ is recovered from l_k as

$$\widehat{var}(\hat{\theta}_k) = \frac{(\hat{\theta}_k - l_k)^2}{V^2_{\zeta/2}}.$$

Similarly, the upper limit is obtained. Then, the $100(1-\zeta)\%$ MOVER confidence interval for $\theta_1+\theta_2$ is defined as

$$CI_{\theta_1+\theta_2}=[L_{\theta_1+\theta_2},U_{\theta_1+\theta_2}],$$

where

$$L_{\theta_1+\theta_2}=(\hat{\theta}_1+\hat{\theta}_2)-\sqrt{(\hat{\theta}_1-l_1)^2+(\hat{\theta}_2-l_2)^2}$$

$$U_{\theta_1+\theta_2}=(\hat{\theta}_1+\hat{\theta}_2)+\sqrt{(u_1-\hat{\theta}_1)^2+(u_2-\hat{\theta}_2)^2}.$$

Now, we focus on CIs for ϑ so that we take log-transformation as

$$\ln\vartheta=\ln(1-\delta)+\mu+\frac{\sigma^2}{2}=\eta_1+\eta_2+\eta_3. \tag{2}$$

By substitution $\hat{\mu},\hat{\sigma}^2,\hat{\delta}$ in its parameter, the estimate $\hat{\eta}_1+\hat{\eta}_2+\hat{\eta}_3=\ln(1-\hat{\delta})+\hat{\mu}+\hat{\sigma}^2$ becomes the estimate of $\eta_1+\eta_2+\eta_3$. For considering CI for σ^2, the unbiased estimate of σ^2 is

$$\hat{\sigma}^2=\frac{1}{n_{(1)}-1}\sum_{i=1}^{n_{(1)}}(X_i-\hat{\mu})^2,$$

which is transformed as

$$\frac{(n_{(1)}-1)\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{n_{(1)}-1},$$

where $\hat{\sigma}^2$ denoted as the sample variance for the log-transformed data of non-zeros. $\chi^2_{n_{(1)}-1}$ stands for chi-square distribution with $n_{(1)}-1$ degrees of freedom. At significant level ζ, the coverage probability for $\chi^2_{n_{(1)}-1}$ is $P\left(\chi^2_{\frac{\zeta}{2},n_{(1)}-1}\le\chi^2_{n_{(1)}-1}\le\chi^2_{1-\frac{\zeta}{2},n_{(1)}-1}\right)=1-\zeta$. Therefore, the $100(1-\zeta)\%$ CI for σ^2 is given as

$$CI_{\sigma^2}=[l_{\sigma^2},u_{\sigma^2}]=\left[\frac{(n_{(1)}-1)\hat{\sigma}^2}{\chi^2_{1-\frac{\zeta}{2},n_{(1)}-1}},\frac{(n_{(1)}-1)\hat{\sigma}^2}{\chi^2_{\frac{\zeta}{2},n_{(1)}-1}}\right].$$

Second, the CI for μ is constructed. Here $Y\sim N(\mu,\sigma^2)$ and also $\hat{\mu}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$. By CLT, the random variable Z is

$$W=\frac{\hat{\mu}-\mu}{\sqrt{\frac{\sigma^2}{n_{(1)}}}}=\frac{\hat{\mu}-\mu}{\sqrt{\frac{(n_{(1)}-1)\hat{\sigma}^2}{n_{(1)}U}}}\sim N(0,1).$$

Hence, the $100(1-\zeta)\%$ CI for μ is

$$CI_\mu=[l_\mu,u_\mu]=\left[\hat{\mu}\pm W_{1-\zeta/2}\sqrt{\frac{(n_{(1)}-1)\hat{\sigma}^2}{n_{(1)}U}}\right],$$

where W_ζ be the ζ^{th} of $N(0,1)$. The MOVER confidence interval for ϑ is constructed. From (2), the $100(1-\zeta)\%$ MOVER confidence interval for $\eta_2+\eta_3$ is given by

$$CI_{\eta_2+\eta_3} = [L_{\eta_2+\eta_3}, U_{\eta_2+\eta_3}],$$

where

$$L_{\eta_2+\eta_3} = (\hat{\eta}_2 + \hat{\eta}_3) - \sqrt{(\hat{\eta}_2 - l_\mu)^2 + \left(\hat{\eta}_3 - \frac{l_{\sigma^2}}{2}\right)^2} \text{ and } U_{\eta_2+\eta_3} = (\hat{\eta}_2 + \hat{\eta}_3) + \sqrt{(u_\mu - \hat{\eta}_2)^2 + \left(\frac{u_{\sigma^2}}{2} - \hat{\eta}_3\right)^2}.$$

The previous step is combined so that the $100(1-\zeta)\%$ MOVER confidence interval for ϑ is given by

$$CI_\vartheta = [L_\vartheta, U_\vartheta],$$

where

$$L_\vartheta = \exp\left[(\hat{\eta}_1 + \hat{\eta}_2 + \hat{\eta}_3) - \sqrt{(\hat{\eta}_1 - l_1)^2 + ((\hat{\eta}_2 + \hat{\eta}_3) - L_{\eta_2+\eta_3})^2}\right] \text{ and}$$

$$U_\vartheta = \exp\left[(\hat{\eta}_1 + \hat{\eta}_2 + \hat{\eta}_3) + \sqrt{(u_1 - \hat{\eta}_1)^2 + (U_{\eta_2+\eta_3} - (\hat{\eta}_2 + \hat{\eta}_3))^2}\right].$$

For l_1 and u_1 derive from three methods for establish CIs for δ: the VST, Wilson score and Jeffreys confidence intervals, as can be seen below.

1) MOVER-VST

The VST for δ was presented by DasGupta (2008), Wu and Hsieh (2014). Recall that $n_{(0)} \sim B(n,\delta)$. The expected Fisher' s information for δ is $I_n(\delta) = n/[\delta(1-\delta)]$. Using delta theorem, we obtain that $\sqrt{n}(\hat{\delta} - \delta) \sim N(0, \delta(1-\delta))$ where $var(\delta) = I_n^{-1}$. The VST is defined as

$$g(\delta) = \int \frac{1/2}{\sqrt{(\delta(1-\delta))}} d\delta = \arcsin(\sqrt{\delta}).$$

Thus, $g(n_{(0)}) = \arcsin\left(\sqrt{\frac{n_{(0)}}{n}}\right)$ and get that $W = 2\sqrt{n}(\arcsin(\hat{\delta}) - \arcsin(\delta)) \sim N(0,1)$ as $n \to \infty$. Then, the $100(1-\zeta)\%$ asymptotically CI for δ is

$$CI_{\delta.v} = [l_{1.v}, u_{1.v}] = \left[\sin^2\left(\arcsin\sqrt{\hat{\delta}} \pm W_{(1-\frac{\zeta}{2})} \frac{1}{2\sqrt{n}}\right)\right].$$

Then, $100(1-\zeta)\%$ MOVER-VST confidence interval for ϑ is given by

$$CI_{m.v} = [L_{m.v}, U_{m.v}],$$

where

$$L_{m.v} = \exp\left[(\hat{\eta}_1 + \hat{\eta}_2 + \hat{\eta}_3) - \sqrt{(\hat{\eta}_1 - l_{1.v})^2 + ((\hat{\eta}_2 + \hat{\eta}_3) - L_{\eta_2+\eta_3})^2}\right] \text{ and}$$

$$U_{m.v} = \exp\left[(\hat{\eta}_1 + \hat{\eta}_2 + \hat{\eta}_3) + \sqrt{(u_{1.v} - \hat{\eta}_1)^2 + (U_{\eta_2+\eta_3} - (\hat{\eta}_2 + \hat{\eta}_3))^2}\right].$$

2) MOVER-Wilson

The Wilson interval for δ was improved by Wilson (1927). After that Wilk (1938) proposed by the score method, called Wilson score method. Donner and Zou (2011) found that it can perform well for small to moderate sample sizes. Thus, the $100(1-\zeta)\%$ Wilson CI for δ is

$$CI_{\delta.w} = [l_{1.w}, u_{1.w}] = \left[\frac{n_{(0)} + T_{\frac{\zeta}{2}}^2/2}{n + T_{\frac{\zeta}{2}}^2} \pm \left(\frac{T_{1-\frac{\zeta}{2}}}{n + T_{\frac{\zeta}{2}}^2} \sqrt{\frac{n_{(0)}n_{(1)}}{n} + \frac{T_{\frac{\zeta}{2}}^2}{4}} \right) \right],$$

where $T = (n_{(0)} - n\delta) / \sqrt{n\delta(1-\delta)} \sim N(0,1)$. Then, $100(1-\zeta)\%$ MOVER-Wilson confidence interval for ϑ is given by

$$CI_{m.w} = [L_{m.w}, U_{m.w}],$$

where

$$L_{m.w} = \exp\left[(\hat{\eta}_1 + \hat{\eta}_2 + \hat{\eta}_3) - \sqrt{(\hat{\eta}_1 - l_{1.w})^2 + ((\hat{\eta}_2 + \hat{\eta}_3) - L_{\eta_2+\eta_3})^2} \right] \text{ and}$$

$$U_{m.w} = \exp\left[(\hat{\eta}_1 + \hat{\eta}_2 + \hat{\eta}_3) + \sqrt{(u_{1.w} - \hat{\eta}_1)^2 + (U_{\eta_2+\eta_3} - (\hat{\eta}_2 + \hat{\eta}_3))^2} \right].$$

3) MOVER-Jeffreys

The Jeffreys method was developed from Brown et al. (2001) for using beta prior in inference on δ (Berger 1985). Let $\text{beta}(b_1, b_2)$ and $\text{beta}(n_{(0)} + b_1, n_{(1)} + b_2)$ be the prior and posterior distributions of δ, respectively such that Jeffreys prior has the distribution $\text{beta}\left(\frac{1}{2}, \frac{1}{2}\right)$. The $100(1-\zeta)\%$ Jeffreys CI for δ is

$$CI_{\delta.J} = [l_{1.J}, u_{1.J}] = \left[\text{beta}\left(\frac{\zeta}{2}; \alpha, \beta \right), \text{beta}\left(1 - \frac{\zeta}{2}; \alpha, \beta \right) \right],$$

where $\alpha = n_{(0)} + 1/2$ and $\beta = n_{(1)} + 1/2$. Then, $100(1-\zeta)\%$ MOVER-Jeffreys confidence interval for ϑ is given by

$$CI_{m.j} = [L_{m.j}, U_{m.j}],$$

where

$$L_{m.J} = \exp\left[(\hat{\eta}_1 + \hat{\eta}_2 + \hat{\eta}_3) - \sqrt{(\hat{\eta}_1 - l_{1.J})^2 + ((\hat{\eta}_2 + \hat{\eta}_3) - L_{\eta_2+\eta_3})^2} \right] \text{ and}$$

$$U_{m.J} = \exp\left[(\hat{\eta}_1 + \hat{\eta}_2 + \hat{\eta}_3) + \sqrt{(u_{1.J} - \hat{\eta}_1)^2 + (U_{\eta_2+\eta_3} - (\hat{\eta}_2 + \hat{\eta}_3))^2} \right].$$

3. Methods for the Difference between Two Delta-Lognormal Means

The CIs for the difference between two means are expanded from the previous section using the concepts of MOVER based on VST, Wilson score and Jeffreys, compared with GCI. Let $X = (X_1, X_2, ..., X_n)$ and $V = (V_1, V_2, ..., V_m)$ be two non-negative and independent random variables of delta-lognormal distribution, denoted as $\Delta(\mu, \sigma^2, \delta)$ and $\Delta(\mu_2, \sigma_2^2, \delta_2)$, respectively. The maximum likelihood estimators for μ_2, σ_2^2 and δ_2 are $\hat{\mu}_2 = \frac{1}{m_{(1)}} \sum_{j=1}^{m_{(1)}} \ln v_j$, $\hat{\sigma}_2^2 = \frac{1}{m_{(1)}} \sum_{j=1}^{m_{(1)}} \left(\ln v_j - \hat{\mu}_2 \right)^2$ and $\hat{\delta}_2 = \frac{m_{(0)}}{m}$; $m_{(0)} + m_{(1)} = m$, respectively. Then, the difference between two delta-lognaormal means is given by

$$\xi = \vartheta - \vartheta_2,$$

where $\vartheta_2 = (1-\delta_2)\exp\left(\mu_2 + \frac{\sigma_2^2}{2}\right)$. The $\hat{\xi} = \hat{\vartheta} - \hat{\vartheta}_2$ is the point estimate of ξ.

3.1. GCI for the difference between two means

The GCI conditions are detailed in Section 2. The importance is the GPQ for ξ used to establish GCI. Since that ξ is the function of parameters μ_2, σ_2^2 and δ_2. By random variable V, the GPQs for μ_2, σ_2^2 and δ_2 are defined as

$$R_{\mu_2} = \hat{\mu}_2 - \frac{W'}{\sqrt{m_{(1)}}}\sqrt{\frac{(m_{(1)}-1)\hat{\sigma}_2^2}{U'}},$$

$$R_{\sigma_2^2} = \frac{(m_{(1)}-1)\hat{\sigma}_2^2}{U'},$$

$$R_{\delta_2} = \sin^2\left[\arcsin(\hat{\delta}_2) - \frac{Z'}{2\sqrt{m}}\right],$$

where $W' = (\hat{\mu}_2 - \mu_2)\Big/\sqrt{\frac{(m_{(1)}-1)\hat{\sigma}_2^2}{m_{(1)}U_2}} \sim N(0,1)$, $\quad Z' = 2\sqrt{m}(\arcsin\sqrt{\hat{\delta}_2} - \arcsin\sqrt{\delta_2}) \sim N(0,1)$ and $U' \sim \chi^2_{m_{(1)}-1}$. By the pivots of $R_\mu, R_{\mu_2}, R_{\sigma^2}, R_{\sigma_2^2}, R_\delta, R_{\delta_2}$, the equivalent of ξ is given by

$$\xi = (1-R_\delta)\exp\left(R_\mu + \frac{R_{\sigma^2}}{2}\right) - (1-R_{\delta_2})\exp\left(R_{\mu_2} + \frac{R_{\sigma_2^2}}{2}\right).$$

It can be written as

$$R_\xi = \left(1-\sin^2\left[\arcsin(\hat{\delta}) - \frac{Z}{2\sqrt{n}}\right]\right)\exp\left[\hat{\mu} - W\sqrt{\frac{n_{(1)}-1)\hat{\sigma}^2}{n_{(1)}U}} + \frac{(n_{(1)}-1)\hat{\sigma}^2}{2U}\right]$$
$$-\left(1-\sin^2\left[\arcsin(\hat{\delta}_2) - \frac{Z'}{2\sqrt{m}}\right]\right)\exp\left[\hat{\mu}_2 - W'\sqrt{\frac{m_{(1)}-1)\hat{\sigma}_2^2}{m_{(1)}U'}} + \frac{(m_{(1)}-1)\hat{\sigma}_2^2}{2U'}\right], \quad (3)$$

where the random variables U, U', W, W', Z and Z' are independent. The GPQ in (3) satisfies the two conditions for being a GPQ. As a result, the $100(1-\zeta)\%$ GCI for ξ is

$$CI_{dgci} = [L_{dgci}, U_{dgci}] = [R_\xi(\zeta/2), R_\xi(1-\zeta/2)],$$

where $R_\xi(\zeta)$ denotes the ζ^{th} percentile of R_ξ.

3.2. MOVERs for the difference between two means

The MOVER idea based on Zou et al. (2009a) was described to construct the CIs for $c_1\theta_1 + c_2\theta_2$; c_1 and c_2 are constants, defined as

$$CI'_{dm} = [L'_{dm}, U'_{dm}],$$

where

$$L'_{dm} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2 - \sqrt{\left[c_1\hat{\theta}_1 - \min(c_1l_1, c_1u_1)\right]^2 + \left[c_2\hat{\theta}_2 - \min(c_2l_2, c_2u_2)\right]^2} \text{ and}$$

$$U'_{dm} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2 + \sqrt{\left[c_1\hat{\theta}_1 - \max(c_1l_1, c_1u_1)\right]^2 + [c_2\hat{\theta}_2 - \max(c_2l_2, c_2u_2)]^2}.$$

This leads to construct MOVER confidence interval for $\xi_1 - \xi_2$, given by

$$CI_{dm} = [L_{dm}, U_{dm}],$$

where

$$L_{dm} = (\hat{\vartheta} - \hat{\vartheta}_2) - \sqrt{[\hat{\vartheta} - \min(l_1, u_1)]^2 + [-\hat{\vartheta}_2 - \min(-l_2, -u_2)]^2} \text{ and}$$

$$U_{dm} = (\hat{\vartheta} - \hat{\vartheta}_2) + \sqrt{[\hat{\vartheta} - \max(l_1, u_1)]^2 + [-\hat{\vartheta}_2 - \max(-l_2, -u_2)]^2}.$$

For l_k and u_k, $k = 1,2$ are depend on the CIs for ϑ and ϑ_2, meanwhile CIs for δ consist of three methods: VST, Wilson score and Jeffreys elaborated as the previous section.

1) MOVER-VST

CIs for μ, σ^2 and δ were presented in Section 2. Furthermore, setting

$$l_{1v} = (1 - u_{1.v}) \exp\left(l_\mu + \frac{l_{\sigma^2}}{2} \right),$$

$$u_{1v} = (1 - l_{1.v}) \exp\left(u_\mu + \frac{u_{\sigma^2}}{2} \right),$$

and

$$l_{2v} = (1 - u_{2.v}) \exp\left(l_{\mu_2} + \frac{l_{\sigma_2^2}}{2} \right),$$

$$u_{2v} = (1 - l_{2.v}) \exp\left(l_{\mu_2} + \frac{u_{\sigma_2^2}}{2} \right),$$

where

$$CI_{\mu_2} = [l_{\mu_2}, u_{\mu_2}] = \left[\hat{\mu}_2 \pm W' \sqrt{\frac{(m_{(1)} - 1)\hat{\sigma}^2}{m_{(1)} U'}} \right],$$

$$CI_{\sigma_2^2} = [l_{\sigma_2^2}, u_{\sigma_2^2}] = \left[\frac{(m_{(1)} - 1)\hat{\sigma}_2^2}{\chi^2_{1-\frac{\zeta}{2}, m_{(1)}-1}}, \frac{(m_{(1)} - 1)\hat{\sigma}_2^2}{\chi^2_{\frac{\zeta}{2}, m_{(1)}-1}} \right],$$

$$CI_{\delta.2.v} = [l_{2.v}, u_{2.v}] = \left[\sin^2\left(\arcsin\sqrt{\hat{\delta}_2} \mp Z_{(1-\frac{\zeta}{2})} \frac{1}{2\sqrt{m}} \right) \right],$$

$$U' \sim \chi^2_{m_{(1)}-1},$$

$$W' = (\hat{\mu}_2 - \mu_2) \Big/ \sqrt{\frac{(m_{(1)} - 1)\hat{\sigma}_2^2}{m_{(1)} U_2}} \sim N(0,1),$$

$$Z = 2\sqrt{m}\left(\arcsin(\sqrt{\hat{\delta}_2}) - \arcsin(\sqrt{\delta_2}) \right) \sim N(0,1).$$

Then, the $100(1-\zeta)\%$ MOVER-VST confidence intervals for ξ is given by

$$CI_{dm.v} = [L_{dm.v}, U_{dm.v}],$$

where

$$L_{dm.v} = (\hat{\vartheta} - \hat{\vartheta}_2) - \sqrt{\left[\hat{\vartheta} - \min(l_{1v}, u_{1v})\right]^2 + \left[-\hat{\vartheta}_2 - \min(-l_{2v}, -u_{2v})\right]^2} \text{ and}$$

$$U_{dm.v} = (\hat{\vartheta} - \hat{\vartheta}_2) + \sqrt{\left[\hat{\vartheta} - \max(l_{1v}, u_{1v})\right]^2 + \left[-\hat{\vartheta}_2 - \max(-l_{2v}, -u_{2v})\right]^2}.$$

2) MOVER-Wilson

Similarly, given

$$l_{1w} = (1 - u_{1.w})\exp\left(l_\mu + \frac{l_{\sigma^2}}{2}\right),$$
$$u_{1w} = (1 - l_{1.w})\exp\left(l_\mu + \frac{u_{\sigma^2}}{2}\right),$$

and

$$l_{2w} = (1 - u_{2.w})\exp\left(l_{\mu_2} + \frac{l_{\sigma_2^2}}{2}\right),$$
$$u_{2w} = (1 - l_{2.w})\exp\left(l_{\mu_2} + \frac{u_{\sigma_2^2}}{2}\right),$$

where

$$CI_{\delta.2.w} = [l_{2.w}, u_{2.w}] = \left[\frac{m_{(0)} + \tilde{T}^2_{\frac{\zeta}{2}}/2}{m + \tilde{T}^2_{\frac{\zeta}{2}}} \mp \left(\frac{\tilde{T}_{1-\frac{\zeta}{2}}}{m + \tilde{T}^2_{\frac{\zeta}{2}}}\sqrt{\frac{m_{(0)}m_{(1)}}{m} + \frac{\tilde{T}^2_{\frac{\zeta}{2}}}{4}}\right)\right] \text{ and}$$

$T = (m_{(0)} - m\delta)/\sqrt{m\delta(1-\delta)} \sim N(0,1)$.

Hence, the $100(1-\zeta)\%$ MOVER-Wilson confidence interval for ξ is

$$CI_{dm.w} = [L_{dm.w}, U_{dm.w}],$$

where

$$L_{dm.w} = (\hat{\vartheta} - \hat{\vartheta}_2) - \sqrt{\left[\hat{\vartheta} - \min(l_{1w}, u_{1w})\right]^2 + \left[-\hat{\vartheta}_2 - \min(-l_{2w}, -u_{2w})\right]^2} \text{ and}$$
$$U_{dm.w} = (\hat{\vartheta} - \hat{\vartheta}_2) + \sqrt{\left[\hat{\vartheta} - \max(l_{1w}, u_{1w})\right]^2 + \left[-\hat{\vartheta}_2 - \max(-l_{2w}, -u_{2w})\right]^2}.$$

3) MOVER-Jeffreys

Also, let

$$l_{1J} = (1 - u_{1.J})\exp\left(l_\mu + \frac{l_{\sigma^2}}{2}\right),$$
$$u_{1J} = (1 - l_{1.J})\exp\left(l_\mu + \frac{u_{\sigma^2}}{2}\right),$$

and

$$l_{2J} = (1 - u_{2.J})\exp\left(l_{\mu_2} + \frac{l_{\sigma_2^2}}{2}\right),$$
$$u_{2J} = (1 - l_{2.J})\exp\left(l_{\mu_2} + \frac{u_{\sigma_2^2}}{2}\right),$$

where

$$CI_{\delta.2.J}=[l_{2.J},u_{2.J}]=\left[\text{beta}\left(\frac{\zeta}{2};\alpha_2,\beta_2\right),\text{beta}\left(1-\frac{\zeta}{2};\alpha_2,\beta_2\right)\right],$$

$\alpha_2=m_{(0)}+1/2$ and $\beta_2=m_{(1)}+1/2.$

Therefore, the $100(1-\zeta)\%$ MOVER-Jeffreys confidence interval for ξ is

$$CI_{dm.J}=[L_{dm.J},U_{dm.J}],$$

where

$$L_{dm.J}=(\hat{\vartheta}-\hat{\vartheta}_2)-\sqrt{\left[(\hat{\vartheta}-\min(l_{1J},u_{1J})\right]^2+\left[-\hat{\vartheta}_2-\min(-l_{2J},-u_{2J})\right]^2}\ \text{ and}$$

$$U_{dm.J}=(\hat{\vartheta}-\hat{\vartheta}_2)+\sqrt{\left[(\hat{\vartheta}-\max(l_{1J},u_{1J})\right]^2+\left[-\hat{\vartheta}_2-\max(-l_{2J},-u_{2J})\right]^2}.$$

4. Simulation Studies

This section shows that Monte Carlo simulation is proceeded using R statistical programming (Venables et al. 2015) to examine the proposed CI performances in terms of coverage probability (CP) and average length (AL). At the nominal confidence level 0.95, all methods are used 10,000 replications and 5,000 GPQs for the GCI. In general, there are two important properties to find the recommended CI: CP with close or greater than a nominal coverage level and providing the shortest AL.

For focusing on CIs for delta-lognormal mean, the random samples are generated from $\Delta(\mu,\sigma^2,\delta)$. The mean μ set to 0. All CIs for ϑ are the comparison between sample size $n=20,50,100$, the probability of having zero $\delta=0.2,0.5,0.8$ and the coefficient of variation $\phi=0.2,0.5,1.0,2.0$. In this study, the cases of $n=20,\delta=0.8$ and $\phi=0.2,0.5,1.0,2.0$ are excluded because the number of positive value $E(n_{(1)})$ is less than 10. Fletcher (2008), Wu and Hsieh (2014) claimed that these combination affecting the CI performances. To construct CIs for ϑ, we propose the MOVER-VST, MOVER-Wilson and MOVER-Jeffreys which are compared with the existing GCI. Table 1, the numerical results revealed that the MOVER-VST and MOVER-Wilson provided their CPs over the nominal level excluding cases of $\phi=2$ and small sample sizes. Both were also excepted for $\delta=0.8$, $\phi=2$ and large sample sizes. On the contrary, the MOVER-Jeffreys performance gave the CPs being less than 0.95 for $\phi=2$. The CP performance of GCI tended to close a desired value, although its ALs were wider than other methods for $\phi=2$ and small sample sizes.

For examining on CIs for the difference between two delta-lognormal means, the random samples are generated from $\Delta(\mu,\sigma^2,\delta)$ and $\Delta(\mu_2,\sigma_2^2,\delta_2)$ to establish CIs for ξ. All parameters is relative with $\Delta(\mu_2,\sigma_2^2,\delta_2)$ which are similarly fixed as mentioned above. Table 2 showed that the CP from three MOVER were closer or greater than a given target almost all cases, however the situations of $\delta=\delta_2=0.2$ and $\phi=\phi_2=1,2$ were omitted. Importantly, MOVER-Jeffreys became the lowest AL. The GCI can maintain the target for the rest cases.

Table 1 The CPs and AL performances of 95% CIs for ϑ

n	δ	ϕ	CP				AL			
			MOVER-V	MOVER-W	MOVER-J	GCI	MOVER-V	MOVER-W	MOVER-J	GCI
20	0.2	0.2	0.941	0.981	0.922	0.935	0.437	0.462	0.399	0.355
		0.5	0.959	0.976	0.952	0.949	0.822	0.847	0.802	0.639
		1.0	0.956	0.965	0.953	0.946	1.811	1.831	1.799	1.730
		2.0	0.941	0.945	0.939	0.944	3.661	3.676	3.654	8.337
	0.5	0.2	0.957	0.962	0.900	0.954	0.840	0.805	0.711	0.457
		0.5	0.957	0.961	0.924	0.951	1.277	1.252	1.184	0.710
		1.0	0.960	0.960	0.944	0.952	2.568	2.551	2.505	2.092
		2.0	0.939	0.938	0.930	0.942	5.042	5.030	4.997	39.703
50	0.2	0.2	0.950	0.970	0.939	0.947	0.283	0.290	0.269	0.232
		0.5	0.957	0.968	0.954	0.945	0.519	0.527	0.512	0.378
		1.0	0.970	0.974	0.970	0.946	1.150	1.157	1.148	0.850
		2.0	0.974	0.975	0.974	0.946	2.372	2.376	2.371	2.704
	0.5	0.2	0.955	0.956	0.927	0.953	0.538	0.528	0.495	0.291
		0.5	0.952	0.954	0.935	0.947	0.766	0.759	0.737	0.403
		1.0	0.957	0.958	0.950	0.946	1.472	1.468	1.453	0.823
		2.0	0.949	0.949	0.945	0.943	2.870	2.867	2.857	2.754
	0.8	0.2	0.962	0.945	0.925	0.961	1.067	1.013	0.960	0.248
		0.5	0.967	0.953	0.942	0.961	1.427	1.373	1.339	0.344
		1.0	0.958	0.946	0.936	0.951	2.613	2.562	2.544	0.888
		2.0	0.943	0.934	0.932	0.944	4.994	4.951	4.940	8.484
100	0.2	0.2	0.952	0.967	0.948	0.949	0.206	0.209	0.201	0.167
		0.5	0.968	0.973	0.966	0.949	0.393	0.396	0.391	0.263
		1.0	0.992	0.992	0.992	0.949	0.923	0.926	0.923	0.563
		2.0	0.999	0.999	0.999	0.946	1.968	1.969	1.968	1.610
	0.5	0.2	0.957	0.958	0.943	0.957	0.388	0.384	0.372	0.207
		0.5	0.962	0.963	0.954	0.955	0.556	0.553	0.545	0.276
		1.0	0.971	0.971	0.968	0.947	1.100	1.099	1.094	0.522
		2.0	0.987	0.987	0.986	0.947	2.212	2.211	2.208	1.460
	0.8	0.2	0.959	0.945	0.935	0.957	0.758	0.738	0.718	0.172
		0.5	0.961	0.949	0.945	0.957	0.971	0.950	0.937	0.223
		1.0	0.957	0.948	0.945	0.950	1.704	1.683	1.677	0.429
		2.0	0.945	0.939	0.939	0.946	3.206	3.188	3.185	1.474

Notes: MOVER-V, MOVER-W and MOVER-J denote the MOVER-VST, MOVER-Wilson, MOVER-Jeffreys, respectively.

Table 2 The CPs and AL performances of 95% CIs for ξ

n	m	δ	δ_2	ϕ	ϕ_2	CP				AL			
						MOVER-V	MOVER-W	MOVER-J	GCI	MOVER-V	MOVER-W	MOVER-J	GCI
20	20	0.2	0.2	0.2	0.2	0.969	0.973	0.971	0.983	0.473	0.475	0.443	0.505
				0.5	0.5	0.946	0.948	0.945	0.975	0.779	0.762	0.729	0.921
				1.0	1.0	0.932	0.930	0.929	0.965	2.298	2.234	2.191	2.676
				2.0	2.0	0.938	0.937	0.937	0.959	12.995	12.692	12.566	14.437
		0.5	0.5	0.2	0.2	0.994	0.993	0.990	0.994	0.651	0.624	0.598	0.648
				0.5	0.5	0.990	0.988	0.986	0.989	1.076	1.036	0.988	1.040
				1.0	1.0	0.979	0.977	0.975	0.978	3.588	3.491	3.354	3.385
				2.0	2.0	0.968	0.965	0.964	0.964	48.833	47.876	46.334	46.743
50	50	0.2	0.2	0.2	0.2	0.982	0.982	0.981	0.987	0.313	0.310	0.299	0.329
				0.5	0.5	0.947	0.945	0.945	0.972	0.468	0.459	0.448	0.538
				1.0	1.0	0.927	0.925	0.924	0.961	1.073	1.050	1.038	1.245
				2.0	2.0	0.931	0.929	0.929	0.955	3.729	3.662	3.644	4.144
		0.5	0.5	0.2	0.2	0.995	0.994	0.993	0.995	0.413	0.405	0.396	0.412
				0.5	0.5	0.989	0.989	0.986	0.989	0.578	0.567	0.554	0.573
				1.0	1.0	0.975	0.974	0.972	0.974	1.250	1.232	1.206	1.222
				2.0	2.0	0.964	0.964	0.962	0.963	4.641	4.593	4.518	4.477
		0.8	0.8	0.2	0.2	0.997	0.997	0.997	0.997	0.362	0.369	0.352	0.354
				0.5	0.5	0.996	0.996	0.995	0.994	0.561	0.575	0.545	0.505
				1.0	1.0	0.992	0.992	0.991	0.985	1.714	1.758	1.673	1.440
				2.0	2.0	0.982	0.981	0.98	0.970	23.29	23.886	22.993	19.172
100	100	0.2	0.2	0.2	0.2	0.990	0.988	0.987	0.993	0.227	0.224	0.220	0.236
				0.5	0.5	0.957	0.955	0.955	0.979	0.329	0.324	0.32	0.373
				1.0	1.0	0.929	0.928	0.927	0.961	0.700	0.689	0.685	0.808
				2.0	2.0	0.923	0.922	0.922	0.952	2.145	2.119	2.113	2.389
		0.5	0.5	0.2	0.2	0.994	0.994	0.994	0.994	0.293	0.290	0.286	0.293
				0.5	0.5	0.989	0.989	0.988	0.989	0.393	0.389	0.385	0.392
				1.0	1.0	0.975	0.975	0.974	0.975	0.763	0.758	0.750	0.756
				2.0	2.0	0.960	0.959	0.958	0.960	2.257	2.245	2.227	2.210
		0.8	0.8	0.2	0.2	0.998	0.998	0.997	0.997	0.248	0.252	0.246	0.245
				0.5	0.5	0.996	0.996	0.995	0.993	0.34	0.346	0.337	0.320
				1.0	1.0	0.989	0.990	0.989	0.983	0.724	0.738	0.719	0.645
				2.0	2.0	0.980	0.980	0.979	0.971	2.807	2.857	2.789	2.460

Notes: MOVER-V, MOVER-W and MOVER-J denote the MOVER-VST, MOVER-Wilson, MOVER-Jeffreys, respectively.

5. Applications

All CIs were established by the different methods in order to apply with real world examples. There are the following data sets in several fields, including environmental problem, fishery survey and biology as seen below.

Data set 1:

The measured airborne chlorine were collected during a working day at an industrial site, United States. National Research Council (US) Committee on Toxicology had noted that chlorine was produced annually about 10 million tons for industrial use (Research National Council 1984). This is air contaminants to affect human health such as symptomatic, with cough, chest tightness, and shortness of breath for fourteen to sixteen hours after exposure. Owen and DeRouen (1980), Tian (2005) and Tian and Wu (2006) applied this dataset with their study. The sample sizes are 15 numbers of measurements including: 9 positive airborne chlorine and the rest is empty. Table 3 reveals that the airborne chlorine observations measured in parts per million (ppm).

Table 3 Data of airborne chlorine concentrations at US industrial site in the period of a working day

Measurement	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Chlorine reading	6	0	6	9	6.5	0	0	0	1	0.5	2	2	0	0	1

Data set 2:

The National Institute of Water and Atmospheric Research was recorded the red cod (Pseudophycis bachus) density at New Zealand trawls. Fletcher (2008) and Wu and Hsieh (2014) utilized this to illustrate in their works as well. There are 67 trawls measured in kilograms per square kilometer (kg/km^2), including 54 positive density and the remainder is blank trawls. The positive density are shown as Table 4.

At the 5% significant level, we used the normality test being the Shapiro-Wilk test, this had a p-value of 0.2355 for the logged positive airborne chlorine. This dataset is together with zero observations, then this can be summarized as delta-lognormal distribution. The estimated airborne chlorine mean is $\hat{\vartheta}_{airbone} = 2.573$, $\hat{\mu} = 0.927$, $\hat{\sigma}^2 = 1.057$, $\hat{\delta} = 0.400$ and $\hat{\phi} = 1.948$. The data was used to compute the 95% CIs for $\vartheta_{airbone}$, as can be seen in Table 5. This result is consistent with the simulation results of $n = 20, \hat{\delta} = 0.5$ and $\hat{\phi} = 2$ indicating MOVER-Wilson is the best CI performance.

Table 4 Data of positive red cod density in New Zealand

Fish density									
10.8	13.2	18.2	19.6	34.2	37.0	41.5	42.3	46.1	46.3
52.7	53.8	55.5	59.2	64.5	66.0	70.2	70.6	74.7	76.8
77.6	78.8	85.0	88.1	89.9	90.8	95.4	100.9	114.1	123.2
131.8	132.7	135.1	141.4	147.4	183.0	223.0	235.3	246.5	253.5
267.1	276.4	293.7	298.6	465.2	584.2	639.2	639.3	663.3	915.7
1004.2	1402.2	1563.2	2948.8						

This data contain zero observations, meanwhile the non-zero observations are investigated to test the normality. We found that the p-value of Shapiro-Wilk test is 0.2544 for the log-transformed quantities of red cod at the 5% significant level. As a result, it fits for the delta-lognormal distribution. This leads to compute the red cod density mean is $\hat{\vartheta}_{fish} = 219.335$ where $\hat{\mu} = 4.864$, $\hat{\sigma}^2 = 1.485$, $\hat{\delta} = 0.194$ and $\hat{\phi} = 2.116$. We focus on the computation of 95% CIs for ϑ_{fish}, included in Table 5. Furthermore, the CI results applying fish density indicate that the MOVER-Wilson is the best confirming with numerical computation in simulation studies section.

Table 5 The 95% CIs for delta-lognormal mean ϑ based on two datasets: the airborne chlorine records and fish densities

CIs	Methods			
	MOVER-VST	MOVER-Wilson	MOVER-Jeffreys	GCI
Dataset 1: Airborne chlorine				
Lower	1.080	1.147	1.079	1.115
Upper	5.454	5.314	5.426	15.839
Length	4.374	4.167	4.347	14.724
Dataset 2: Red cod density				
Lower	143.193	142.996	142.976	147.69
Upper	332.617	331.544	332.120	377.325
Length	189.424	188.548	189.144	229.635

Data set 3:

The distance traveled of mice is a measure how quickly mice from one place to another in meter (m). In biology, it is the study of the activity patterns of a species of field mice, divided by seasons: fall, winter, spring and summer. In this study, the distance traveled data was selected as fall and summer seasons especially, taken from Koopmans (1981) as shown in Table 6. Histogram plots of mice distance in two seasons are displayed as Figure 1.

Table 6 The distance traveled data of mice during fall and summer seasons

Seasons	Distance traveled (m)													
Fall	0	0	21	0	15	0	15	15	0	8	0	0	15	21
	0	34	0	15	8	29	15	46	39	30	15	11	0	
Summer	60	21	15	15	15	33	24	33	42	54	11	32	8	71
	150	18	12	0	0	21	17	0	15	106	17	21	21	

Table 7 Results of distance traveled of field mice on fall and summer seasons

Fall	Summer
$\hat{\mu} = 2.905$	$\hat{\mu}_2 = 3.245$
$\hat{\sigma}^2 = 0.263$	$\hat{\sigma}_2^2 = 0.534$
$\hat{\delta} = 0.373$	$\hat{\delta}_2 = 0.111$
$n = 27$	$m = 27$
$\hat{\phi} = 1.032$	$\hat{\phi}_2 = 0.959$
$\hat{\vartheta} = 13.118$	$\hat{\vartheta}_2 = 29.854$

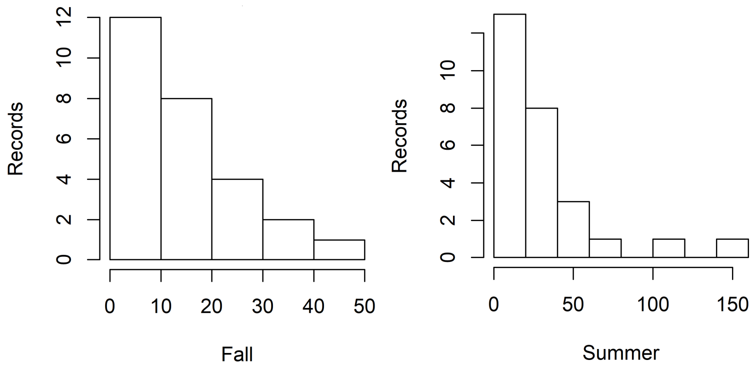

Figure 1 Histogram plots of the distance of mice traveling on seasons: (a) fall and (b) summer

Table 8 The 95% CIs for ξ based on the distance traveled of mice

CIs	Methods			
	MOVER-VST	MOVER-Wilson	MOVER-Jeffreys	GCI
Lower	−21.634	−21.338	−21.087	−33.273
Upper	−10.781	−10.945	−11.653	−5.724
Length	10.853	10.393	9.434	27.549

To investigate the normality, the p-values of Shapiro-Wilk test are 0.1913 and 0.1122 for the log-transformation of positive-valued distances of fall and summer seasons at 5% significant level, respectively. The zero distances are carried in mentioned seasons so that both datasets are considered as the delta-lognormal distribution. Table 7 is displayed the summary statistics. Therefore, the estimated mean difference between summer and fall is $\hat{\zeta}_{mice} = -16.736$ m. Table 8 shows that 95%CIs for ξ. It can be interpreted that the mice behavior with traveled distance during fall season is less than summer period. The mentioned example results show that the best performance is MOVER-Jeffreys following with the numerical results.

6. Discussion and Conclusions

The study aimed to develop statistical estimation as CIs for the single and difference between two means in delta-lognormal distributions. There are three proposed CIs, including the MOVER-VST, MOVER-Wilson and MOVER-Jeffreys. By way of comparison, these CIs were compared with the existing GCI of Wu and Hsieh (2014). The CP and AL performances are used to assess the proposed CIs through Monte Carlo simulation.

For the single mean, the finding can be conclude that the GCI is stable CI in terms of CPs, although its ALs are considered indicating GCI performance is not better than other methods for the large coefficient of variation and small sample size. These GCI results are in agreement with Wu and Hsieh (2014). The MOVER-VST and MOVER-Wilson performances perform well in terms of CP, unless caused by the large coefficient of variation for small sample size and both large zero proportion and coefficient of variation for large sample size. Importantly, it is easier to compute than the GCI.

Therefore, both can be considered as recommended CIs for the mean. On the other side, the CP performances of MOVER-Jeffreys were quite under the target for the small coefficient of variation and all sample sizes, it is then not suggested. According to Donner and Zou (2011), Wilson and Jeffreys CIs for δ were recommended, although MOVER-Jeffreys for delta-lognormal mean is not a good performance. It is possible that Jeffreys CI is directly generated from the probability of having non-zero observation. For the difference between two means, the MOVER-Jeffreys becomes to a recommended CI, meanwhile the cases of small zero proportion and large dispersion are excluded. On the other hand, the GCI is also recommended in the reminder MOVER-Jeffreys cases.

Acknowledgments

The authors are grateful to the reviewers and the editor for their suggestions and comments which helped us to improve the quality of the manuscript. The first author was funded by Thailand Science Research and Innovation (TSRI) and National Research Council of Thailand (NRCT): Contract no. PHD/0198/2561. This was also supported by King Mongkut's University of Technology North Bangkok and Uttaradit Rajabhat University.

References

Aitchison J. On the distribution of a positive random variable having a discrete probability and mass at the origin. J Am Stat Assoc. 1955; 50: 901-908.

Berger JO. Statistical decision theory and Bayesian analysis. New York: Springer-Verlag; 1985.

Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Stat Sci. 2001; 16: 101-133.

Casella G, Berger RL. Statistical Inference. California: Duxbury Press; 2002.

Chen YH, Zhou XH. Generalized confidence intervals for the ratio or difference of two means for lognormal populations with zeros. UW Biostatistics Working Paper Series. Working Paper 296; 2006. p.1-26.

DasGupta A. Asymptotic theory of statistics and probability. New York: Springer; 2008.

Donner A, Zou GY. Estimating simultaneous confidence intervals for multiple contrasts of proportions by the method of variance estimates recovery. Stat Biopharm Res. 2011; 3: 320-335.

Donner A, Zou GY. Closed-form confidence intervals for functions of the normal mean and standard deviation. Stat Methods Med Res. 2012; 21: 347-359.

Fletcher D. Confidence intervals for the mean of the delta-lognormal distribution. Environ Ecol Stat. 2008; 15: 175-189.

Harvey J, Merwe AJ. Bayesian confidence intervals for means and variances of lognormal and bivariate lognormal distributions. J Stat Plan Infer. 2012; 142: 1294-1309.

Hasan MS, Krishnamoorthy K. Confidence intervals for the mean and a percentile based on zero-inflated lognormal data. J Stat Comput Simul. 2018; 88: 1499-1514.

Koopmans L. An introduction to contemporary statistics. Boston: Duxbury Press; 1981.

Krishnamoorthy K, Mathew T. Inferences on the means of lognormal distributions using generalized p-values and generalized confidence intervals. J Stat Plan Infer. 2003; 115: 103-121.

Krishnamoorthy K, Oral E. Standardized likelihood ratio test for comparing several log-normal means and confidence interval for the common mean. Stat Methods Med Res. 2017; 26: 2919-2937.

Li X, Zhou X, Tian L. Interval estimation for the mean of lognormal data with excess zeros. Stat Prob Lett. 2013; 83: 2447-2453.

Lo NC, Jacobson LD, Squire, JL. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can J Fish Aquat Sci. 1992; 49: 2515-2526.

Owen WJ, DeRouen TA. Estimation of the mean for lognormal data containing zeroes and left-censored values, with applications to the measurement of worker exposure to air contaminants. Biometrics. 1980; 36: 707-719.

Pennington M. Efficient estimators of abundance, for fish and plankton surveys. Biometrics. 1983; 39: 281-286.

National Research Council (US) Committee on Toxicology. Emergency and continuous exposure limits for selected airborne contaminants: Volume 2. USA: National Academies Press; 1984. Chlorine. Available from: https://www.ncbi.nlm.nih.gov/books/NBK208300/

Rosales KV, Naranjo JD. Confidence intervals for mean difference between two delta-distributions. In: Liu R, McKean J, editors. Robust rank-based and nonparametric methods. Springer Proceedings in Mathematics and Statistics, vol 168. Springer, Cham; 2016. p. 261-272.

Smith SJ. Evaluating the efficiency of the Δ-distribution mean estimator. Biometrics. 1988; 44: 485-493.

Smith SJ. Use of statistical models for the estimation of abundance from groundfish trawl survey data. Can J Fish Aquat Sci. 1990; 47: 894-903.

Stewart C, Field C. Managing the essential zeros in quantitative fatty acid signature analysis. J Agric Biol Environ Stat. 2011; 16: 45-69.

Tian L. Inferences on the mean of zero-inflated lognormal data: the generalized variable approach. Stat Med. 2005; 24: 3223-3232.

Tian L, Wu J. Confidence intervals for the mean of lognormal data with excess zeros. Biom J. 2006; 48: 149-156.

Tian L, Wu J. Inferences on the common mean of several log-normal populations: The generalized variable approach. Biom J. 2007; 49: 944-951.

Tierney WM, Overhage JM, Murray M, Harris LE, Zhou XH, Eckert GJ, Smith FE, Nienaber N, McDonald CJ, Wolinsky FD. Effects of computerized guidelines for managing heart disease in primary care. J Gen Intern Med. 2003; 18: 967-976.

Venables W, Smith D, and Team, RDC. An introduction to R: Notes on R: a programming environment for data analysis and graphics version 3.6.1. Network Theory, 2019. Available from: https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf.

Weerahandi S. Generalized confidence interval. J Am Stat Assoc. 1993; 88: 899-905.

Wilks SS. Shortest average confidence intervals from large samples. Ann Math Stat. 1938; 9: 166-175.

Wilson EB. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927; 22: 209-212.

Wu WH, Hsieh HN. Generalized confidence interval estimation for the mean of delta-lognormal distribution: an application to New Zealand trawl survey data. J Appl Stat. 2014; 41: 1471-1485.

Zhou XH, Tu W. Confidence intervals for the mean of diagnostic test charge data containing zeros. Biometrics. 2000; 56: 1118-1125.

Zhou XH, Gao S, Hui SL. Methods for comparing the means of two independent log-normal samples. Biometrics. 1997; 53: 1129-1135.

Zou GY, Huang W, Zhang X. A note on confidence interval estimation for a linear function of binomial proportions. Comput Stat Data An. 2009a; 53: 1080-1085.

Zou GY, Huo CY, Taleban J. Simple confidence intervals for lognormal means and their differences with environmental applications. Environmetrics. 2009b; 20: 172-180.