Journal of Environmental Statistics
August 2014, Volume 6, Issue 5. http://www.jenvstat.org

Standardized Likelihood Inference for the Mean and Percentiles of a Lognormal Distribution Based on Samples with Multiple Detection Limits

K. Krishnamoorthy
Department of Mathematics
University of Louisiana at Lafayette

Thomas Mathew
Department of Mathematics and Statistics
University of Maryland Baltimore County

Zhao Xu
Department of Mathematics
University of Louisiana at Lafayette

Abstract

This article investigates standardized versions of the signed likelihood ratio test statistic for inference concerning the mean and percentiles of a lognormal distribution based on samples subject to multiple detection limits. The standardized versions considered are due to DiCiccio, Martin and Stern (2001). Computational algorithms are provided and numerical results are given to assess the performance of the proposed methods, and to make comparisons with competing procedures. It is noted that the standardized signed likelihood ratio test statistics provide accurate inference for the above lognormal parameters even for small samples that include non-detects resulting from the presence of multiple detection limits. Furthermore, in the context of hypothesis testing, they are seen to provide comparable or better performance in terms of power, compared to a test based on the generalized inference methodology. The results are illustrated using two examples on environmental applications.

Keywords: lognormal mean, lognormal percentile, parametric bootstrap, signed likelihood ratio test statistic, type I censoring..

1. Introduction

The lognormal distribution plays a crucial role in the analysis of environmental data and exposure data, since the relevant observations are positive, and very often, positively skewed. Even though normal based methods can be obviously applied for analyzing the log-transformed data, the lognormal distribution still presents some unique features; for example, the lognormal mean and variance are functions of both the mean and variance of the log-transformed data. Inference concerning the lognormal mean (and variance) can be challenging, especially when accurate small sample inference is desired. The problems become especially challenging when the samples are subject to detection limits, which will result in Type I left censored data. Based on a Type I censored sample from a normal distribution, different procedures to compute confidence intervals for the mean, variance and the quantiles are investigated in Jeng and Meeker (2000). The authors do note that for the interval estimation of percentiles, the results are not entirely satisfactory when the data are Type I censored; see also the article by Wong and Wu (2000) and the discussion by Doganaksoy and Schmee (2000). It should be noted that inference concerning the lognormal mean is not addressed in these articles. More recently, Krishnamoorthy, Mallick and Mathew (2011) have investigated inference concerning the mean and percentiles of a lognormal distribution based on a sample subject to a single detection limit. Several competing methods were considered and their performance was numerically studied. The procedures investigated included the signed log-likelihood ratio test (SLRT) statistic, a modified signed log-likelihood ratio test (MSLRT) statistic, and *generalized inference*, i.e., methodology based on the generalized p-value and generalized confidence interval. Even though higher order modifications of the SLRT statistic are expected to improve the accuracy of the normal approximation, the MSLRT statistic due to Fraser, Reid and Wu (2011), investigated in Krishnamoorthy, Mallick and Mathew (2011), showed unsatisfactory performance; in fact, the MSLRT statistic often exhibited worse performance compared to the SLRT statistic in terms of coverage probabilities of the resulting confidence intervals for the lognormal mean and percentiles. However, the generalized inference methodology did exhibit satisfactory performance regardless of the sample size and regardless of the proportion of the data below the detection limit. Even in the presence of multiple detection limits, accuracy of the generalized inference methodology has recently been noted in Krishnamoorthy and Xu (2011), once again for inference concerning the lognormal mean and percentiles. In the absence of a detection limit, the generalized inference methodology for inference concerning a lognormal mean (or for comparing two lognormal means) is investigated in Krishnamoorthy and Mathew (2003), and MSLRT statistics for a single lognormal mean is investigated in Wu, Wong, and Jiang (2003) and Wu, Wong, and Wei (2006); the resulting inferences turned out to be highly accurate. In a recent article, Bhaumik et. al. (2013) have considered the problem of testing hypotheses concerning a single lognormal mean, and have compared four tests: based on Student-t, Edgeworth expansion, generalized p-value, and a permutation approach. The authors conclude that among the four tests they have investigated, only the generalized p-value test and the permutation test can be recommended for practical use. The purpose of this article is to apply two higher order modifications of the SLRT statistic for inference concerning the lognormal mean and percentiles, and compare them with methodologies based on generalized inference.

In order to introduce the set up of our investigation, let Y be a random variable following the lognormal distribution, so that $X = \ln Y \sim N(\mu, \sigma^2)$, where μ and σ^2 are, respectively, the mean and variance of the log-transformed quantity X. The lognormal mean is then given by $\exp\left(\mu + \frac{\sigma^2}{2}\right)$, and the pth percentile of the lognormal distribution is given by $\exp(\mu + z_p\sigma)$, where z_p denotes the pth percentile of the standard normal distribution. Thus inferences concerning the lognormal mean and percentiles are equivalent to those concerning $\mu + \frac{\sigma^2}{2}$ and $\mu + z_p\sigma$, respectively. Let y_1, y_2, ..., y_n be a random sample of size n from the lognormal distribution, and let $x_i = \ln y_i$, $i = 1, 2, ..., n$, so that the x_i's form a random sample from $N(\mu, \sigma^2)$. Suppose the data are subject to k detection limits, to be denoted by DL_1,...,DL_k on the log-scale. We assume without loss of generality that $DL_1 < DL_2 < ... < DL_k$. We also assume that n_i measurements have been obtained by the ith laboratory procedure or device with detection limit DL_i and $n = \sum_{i=1}^{k} n_i$. Furthermore, suppose m_i non-detects are below DL_i, and let $m = \sum_{i=1}^{k} m_i$. Here, m_i's are independent binomial random variables with number of trials n_i and "success probability" $p_i = \Phi\left(\frac{DL_i - \mu}{\sigma}\right)$, where $\Phi(x)$ is the standard normal cumulative distribution function. In the present context, p_i's are unknown proportions of non-detects. This is the set up where the generalized inference methodology is investigated in Krishnamoorthy and Xu (2011), and in Krishnamoorthy, Mallick and Mathew (2011), where the latter paper is in the context of a single detection limit.

We shall now provide a brief outline of the higher order procedures that we propose to use in this article. Even though only the MSLRT statistic due to Fraser, Reid and Wu (2011) is considered in Krishnamoorthy, Mallick and Mathew (2011), several higher order versions of the SLRT statistic are available in the literature, depending on the model assumptions. They also differ in terms of accuracy of the normal approximation, and in terms of ease of implementation; see Barndorff-Nielsen (1991), Skovgaard (1996) and DiCiccio, Martin and Stern (2001). In particular, DiCiccio, Martin and Stern (2001) have proposed two simulation based methods to improve upon the accuracy of the normal approximation of the SLRT statistic; in the present investigation, we shall apply these simulation based methods. In order to introduce these, let $(\psi, \boldsymbol{\lambda}')'$ be a vector parameter in a statistical model, where ψ is a scalar parameter of interest. Furthermore, let $l(\psi, \boldsymbol{\lambda})$ denote the log-likelihood function for $(\psi, \boldsymbol{\lambda}')'$ based on observed data. We also denote the MLE of $(\psi, \boldsymbol{\lambda}')'$ by $(\widehat{\psi}, \widehat{\boldsymbol{\lambda}}')'$. Furthermore, for a fixed ψ, the MLE of the nuisance parameter $\boldsymbol{\lambda}$ will be denoted by $\widehat{\boldsymbol{\lambda}}_\psi$. For inference concerning ψ, the signed log-likelihood ratio test (SLRT) statistic, denoted by $R(\psi)$, is given by

$$R(\psi) = \text{sign}(\widehat{\psi} - \psi)\left[2\left\{l(\widehat{\psi}, \widehat{\boldsymbol{\lambda}}) - l(\psi, \widehat{\boldsymbol{\lambda}}_\psi)\right\}\right]^{1/2}, \tag{1}$$

where $\text{sign}(x)$ is $+1$ or -1 depending on whether $x > 0$ or $x < 0$, respectively. In general, it is known that $R(\psi)$ follows a standard normal distribution up to an error of $O(n^{-1/2})$.

Two third-order accurate methods are proposed in DiCiccio, Martin and Stern (2001). For a fixed value of ψ, let the mean and variance of $R(\psi)$ defined above be denoted by $m(\psi, \hat{\boldsymbol{\lambda}}_\psi)$ and $v(\psi, \widehat{\boldsymbol{\lambda}}_\psi)$, respectively, where these quantities are evaluated at $\boldsymbol{\lambda} = \widehat{\boldsymbol{\lambda}}_\psi$. Now define the modified statistic $R_M(\psi)$ by standardizing $R(\psi)$:

$$R_M(\psi) = \frac{R(\psi) - m(\psi, \hat{\boldsymbol{\lambda}}_\psi)}{\sqrt{v(\psi, \widehat{\boldsymbol{\lambda}}_\psi)}}. \tag{2}$$

Then, as noted in DiCiccio, Martin and Stern (2001), $R_M(\psi)$ follows a standard normal distribution up to an error of $O(n^{-3/2})$. Another third-order accurate method consists of computing the cdf of $R(\psi)$ at a point t as $P_{\widehat{\boldsymbol{\lambda}}_\psi}(R(\psi) \le t)$, where the notation implies that the probability is once again evaluated at $\boldsymbol{\lambda} = \widehat{\boldsymbol{\lambda}}_\psi$, for a fixed ψ. As noted in DiCiccio, Martin and Stern (2001), analytic expressions are very often not available for the mean $m(\psi, \hat{\boldsymbol{\lambda}}_\psi)$ and variance $v(\psi, \widehat{\boldsymbol{\lambda}}_\psi)$ of $R(\psi)$. However, these quantities can be easily approximated based on a parametric bootstrap procedure, i.e., based on Monte Carlo simulation of $R(\psi)$ when $\boldsymbol{\lambda} = \widehat{\boldsymbol{\lambda}}_\psi$, for a fixed ψ. Such a Monte Carlo simulation can also be used to calculate the probability $P_{\widehat{\boldsymbol{\lambda}}_\psi}(R(\psi) \le t)$.

In this article, we investigate the aforementioned third order accurate methods for inference concerning the lognormal mean and percentiles based on samples that are subject to multiple detection limits. It turns out that the above higher order modifications do provide highly accurate inference for the lognormal mean as well as percentiles in the presence of multiple detection limits, regardless of the sample size, and regardless of the proportion of non-detects. As noted in DiCiccio, Martin and Stern (2001), a major appeal of the procedures is that once the SLRT statistic is obtained, the proposed modifications are easily implemented via simulation. Details appear in the next section for inference concerning the lognormal mean. We then take up the inference problems concerning lognormal percentiles based on samples subject to multiple detection limits. In order to assess the accuracy of the proposed methods, numerical results are given consisting of type I error probabilities of tests concerning the lognormal mean and percentiles. The proposed methods indeed exhibit excellent performance irrespective of the sample size, number of detection limits, and proportion of the data below the detection limits. Since the generalized inference methodology is also quite accurate, as noted in Krishnamoorthy and Xu (2011), it is natural to compare it with the higher order procedures. In order to facilitate this comparison, simulated powers are reported. For testing hypothesis concerning the lognormal mean, the higher order procedures turned out to have a significant edge in terms of power, compared to the generalized inference procedure. However, the powers are rather similar, for testing hypothesis concerning lognormal percentiles. The overall conclusion is that the higher order procedures investigated in this article provide accurate and easily implementable methodologies for analyzing lognormal samples subject to multiple detection limits.

2. Likelihood based procedures for the lognormal mean

Recall our notation that in a sample of n observations from the lognormal distribution, n_i measurements have been obtained by the ith laboratory procedure or device with its own detection limit, so that $n = \sum_{i=1}^{k} n_i$. Furthermore, among the n_i observations obtained by the ith laboratory procedure or device, m_i of them are non-detects, i.e., they are below the corresponding detection limit, and let $m = \sum_{i=1}^{k} m_i$. Let DL_1, ..., DL_k denote the detection limits on the log-scale, where we assume that $DL_1 < DL_2 < \dots < DL_k$. Without loss of generality, let the detected observations be y_1, y_2, ..., y_{n-m}, and let $x_i = \ln y_i$, $i = 1, 2, ...,$ $n-m$. Define

$$\bar{x}_d = \frac{1}{n-m} \sum_{i=1}^{n-m} x_i \quad \text{and} \quad s_d^2 = \frac{1}{n-m} \sum_{i=1}^{n-m} (x_i - \bar{x}_d)^2. \tag{3}$$

The log-likelihood function, after omitting a constant term, can be written as

$$\sum_{i=1}^{k} m_i \ln \Phi(z_i^*) - \frac{(n-m)}{2} \ln \sigma^2 - \frac{(n-m)(s_d^2 + (\bar{x}_d - \mu)^2)}{2\sigma^2}, \tag{4}$$

where $z_i^* = \frac{DL_i - \mu}{\sigma}$, $i = 1, ..., k$, and $\Phi(.)$ denotes the standard normal cdf. Now let $\widehat{\mu}$ and $\widehat{\sigma}^2$ denote the maximum likelihood estimates of μ and σ^2, respectively, obtained by maximizing the above log-likelihood function. These can be numerically obtained using the bivariate Newton-Raphson iterative method; for details see the Appendix.

2.1. Testing hypothesis concerning the lognormal mean

The parameter of interest is $\psi = \mu + \frac{\sigma^2}{2}$, and consider the problem of testing the null hypothesis $\psi = \psi_0$. The MLE of ψ is $\widehat{\psi} = \widehat{\mu} + \frac{\widehat{\sigma}^2}{2}$. The above log-likelihood can obviously be written as a function of ψ and σ^2, say $l(\psi, \sigma^2)$. For the fixed null value ψ_0 of ψ, let $\widehat{\sigma}^2_{\psi_0}$ denote the constrained MLE of σ^2; that is, $\widehat{\sigma}^2_{\psi_0}$ is the value of σ^2 that maximizes the likelihood function

$$l(\psi_0, \sigma^2) = \sum_{i=1}^{k} m_i \ln \Phi(z_{i\psi_0}) - \frac{(n-m)}{2} \ln \sigma^2 - \frac{(n-m)}{2\sigma^2} \left[s_d^2 + \left(\bar{x}_d - \psi_0 + \frac{\sigma^2}{2} \right)^2 \right], \tag{5}$$

where $z_{i\psi_0} = \frac{DL_i - \psi_0}{\sigma} + \frac{\sigma}{2}$. The computational details are once again given in the Appendix. In terms of these MLEs, the signed LRT statistic, say $R(\psi_0)$, is given by

$$R(\psi_0) = \text{sign}(\widehat{\psi} - \psi_0) \left\{ 2[l(\widehat{\psi}, \widehat{\sigma}^2) - l(\psi_0, \widehat{\sigma}^2_{\psi_0})] \right\}^{1/2}. \tag{6}$$

We shall now consider the two third order modifications due to DiCiccio, Martin and Stern (2001); see equation (2), and the material following this equation. Let

$$R_M(\psi_0) = \frac{R(\psi_0) - m(\psi_0, \hat{\sigma}^2_{\psi_0})}{\sqrt{v(\psi_0, \widehat{\sigma}^2_{\psi_0})}}, \tag{7}$$

where $m(\psi_0, \hat{\sigma}^2_{\psi_0})$ and $v(\psi_0, \widehat{\sigma}^2_{\psi_0})$ are, respectively, the mean and variance of $R(\psi_0)$ when $\sigma^2 = \hat{\sigma}^2_{\psi_0}$. Both $m(\psi_0, \hat{\sigma}^2_{\psi_0})$ and $v(\psi_0, \widehat{\sigma}^2_{\psi_0})$ can be easily approximated based on a parametric bootstrap procedure consisting of Monte Carlo simulation of $R(\psi_0)$ when $\sigma^2 = \widehat{\sigma}^2_{\psi_0}$. The second method suggested in DiCiccio, Martin and Stern (2001) consists of computing tail probabilities concerning $R(\psi_0)$ evaluated at $\sigma^2 = \widehat{\sigma}^2_{\psi_0}$, which can once again be obtained by Monte Carlo simulation. The appropriate tail area clearly provides a p-value for testing the hypothesis $\psi = \psi_0$.

The following algorithm (parametric bootstrap) can be used for the estimation of $m(\psi_0, \hat{\sigma}^2_{\psi_0})$ and $v(\psi_0, \widehat{\sigma}^2_{\psi_0})$:

Algorithm 1

For a given sample of size n with detection limits $DL_1, ..., DL_k$ (on the log scale), let n_i denote the size of the sample analyzed by the ith laboratory, $i = 1, ..., k$, so that $\sum_{i=1}^{k} n_i = n$, where we also note that $n_i \geq m_i$ for each i.

1. Compute the MLEs $\widehat{\mu}$, $\widehat{\sigma}^2$ and, for a given ψ_0, compute the constrained MLE $\widehat{\sigma}^2_{\psi_0}$; we refer to the Appendix for details regarding the computation of these. Set $\widehat{\mu}_0 = \psi_0 - \widehat{\sigma}^2_{\psi_0}/2$.

2. Generate a sample of size n_i from $N(\widehat{\mu}_0, \widehat{\sigma}^2_{\psi_0})$, $i = 1, ..., k$.

3. Discard the observations from the ith sample that are less than DL_i, $i = 1, ..., k$. Let m_i denote the number of observations below DL_i, $i = 1, ..., k$, so that $m = \sum_{i=1}^{k} m_i$.

4. Compute the MLEs $\widehat{\mu}^*$ and $\widehat{\sigma}^{*2}$ by maximizing (4), and the constrained MLE $\widehat{\sigma}^{*2}_{\psi_0}$ by maximizing (5).

5. Compute $R^*(\psi_0) = \text{sign}(\widehat{\psi}^* - \psi_0)\left\{2[l(\widehat{\psi}^*, \widehat{\sigma}^{*2}) - l(\psi_0, \widehat{\sigma}^{*2}_{\psi_0})]\right\}^{1/2}$, where $\widehat{\psi}^* = \widehat{\mu}^* + \widehat{\sigma}^{*2}/2$

6. Repeat steps 1–4 for a large number of times, say, 10,000.

7. The mean and the standard deviation of the 10,000 $R^*(\psi_0)$ are the parametric bootstrap estimates of $m(\psi_0, \hat{\sigma}^2_{\psi_0})$ and $v(\psi_0, \widehat{\sigma}^2_{\psi_0})$, respectively.

Tail probabilities concerning $R(\psi_0)$ can be similarly evaluated at $\sigma^2 = \widehat{\sigma}^2_{\psi_0}$.

Once $m(\psi_0, \hat{\sigma}^2_{\psi_0})$ and $v(\psi_0, \widehat{\sigma}^2_{\psi_0})$ are obtained, $R_M(\psi_0)$ in (7) can be computed, and can be used for testing the hypothesis H_0: $\psi = \psi_0$, by comparing the value of $R_M(\psi_0)$ to the appropriate standard normal percentile. The second option is to compute a p-value by estimating the appropriate tail probability of $R(\psi_0)$, evaluated at $\sigma^2 = \widehat{\sigma}^2_{\psi_0}$. For example, for testing $H_0 : \psi \geq \psi_0$ against the alternative $H_1 : \psi < \psi_0$, the p-value can be computed as $P_{\sigma^2=\widehat{\sigma}^2_{\psi_0}}[R(\psi_0) \geq r_0]$, where r_0 is the observed value of $R(\psi_0)$.

2.2. Numerical Results

We shall now report the estimated Type I error probabilities of three tests for testing hypothesis concerning the lognormal mean based on samples that are subject to one, two or three detection limits. We shall consider the hypotheses $H_0 : \psi \geq \psi_0$ versus $H_1 : \psi < \psi_0$, and the tests considered are: (i) the test based on the SLRT statistic $R(\psi_0)$ in (6), (ii) the test based on the modified statistic $R_M(\psi_0)$ in (7), and (iii) the test based on the p-value computed as $P_{\sigma^2=\widehat{\sigma}^2_{\psi_0}}[R(\psi_0) \geq r_0]$, where r_0 is the observed value of $R(\psi_0)$. In our tables, the tests are denoted as SLRT, SLRT-PB1 and SLRT-PB2, respectively. While preparing the tables, instead of specifying the detection limit, we have specified the proportion of the data that are expected to be below the respective detection limits. This is denoted by p_1 when there is only one detection limits, (p_1, p_2) when there are two detection limits, and (p_1, p_2, p_3) when there are three detection limits. Throughout, we have assumed a null value $\psi_0 = 3$. The type I error probabilities are reported in Table 1 for sample size $n = 10$, 15 and 20, and $\sigma = 1$ and 3; the value of μ is then determined so that $\psi_0 = \mu + \frac{\sigma^2}{2} = 3$. The values reported in the first block (row wise) of Table 1 are type I error rates based on samples with a single detection limit and proportion of non-detects p_1; the values in the second block are for samples which include two detection limits, and those in the third block are for samples with three detection limits. We observe from the table that, for most cases, the type I error rates the SLRT are always larger than the nominal level, whereas those of SLRT-PB1 and SLRT-PB2 are both very close to the nominal level. It is quite clear that the tests SLRT-PB1 and SLRT-PB2 are both satisfactory even when the proportion of non-detects is quite high.

We shall now consider the powers of three competing tests: the tests SLRT-PB1 and SLRT-PB2, as well as the test based on the generalized p-value. In their recent paper, Krishnamoorthy and Xu (2011) have noted the satisfactory performance of the generalized p-value test in the multiple detection limit scenario, even when the proportion of non-detects is quite high. Here we shall not give details of the generalized p-value test; we refer to Krishnamoorthy and Xu (2011). In our power comparison, we have not included the test based on the SLRT statistic, since its type I error performance is not satisfactory. Numerical results on the power are given in Table 2. For sample sizes $n = 15$ and 30, the first part of Table 2 gives the powers for testing $H_0 : \psi \geq 4.5$ versus $H_1 : \psi < 4.5$, for the parameter values $\mu = 0$ and $\sigma = 3$, 2.5, 2 and 1.5. Since we have assumed $\mu = 0$, the value of ψ is 4.5 when $\sigma = 3$; thus the corresponding power values are simply the type I error probabilities. The remaining entries are the powers. The second part of Table 2 assumes $\sigma = 1$, and considers the testing of $H_0 : \psi \geq 3.5$ versus $H_1 : \psi < 3.5$. Powers values are now tabulated for $\mu = 3$, 2.5, 2, 1.5, where the power values when $\mu = 3$ are simply the type I error probabilities. The type I error probabilities are quite satisfactory for the generalized p-value test, as already noted in Krishnamoorthy and Xu (2011). The overall conclusion regarding the power is that both tests SLRT-PB1 and SLRT-PB2 have an edge over the generalized p-value test in terms of power. The gain in power is sometimes rather substantial for SLRT-PB1; for example, see the powers corresponding to $\mu = 2$ in the second part of the table. It should be noted that the generalized p-value test also requires the computation of the MLE, along with simulation, for its implementation, as noted in Krishnamoorthy and Xu (2011). The computational effort required to implement SLRT-PB1 and SLRT-PB2 is essentially the same as that required for the generalized p-value test. The satisfactory performance in terms of type I error, and the improved performance in terms of power make the tests SLRT-PB1 and SLRT-PB2 attractive for practical use. Of particular interest is the observation that the suggested tests exhibit excellent performance even when the sample size is not large, and regardless of the proportion of the data below the detection limits. Our limited simulations results suggest that the test SLRT-PB1 is to be preferred among the three tests.

Clearly, confidence intervals for the lognormal mean can be obtained using the above test statistics. Furthermore, the confidence intervals will have coverage probabilities close to the nominal level, except those based on the SLRT statistic, and these are expected to be liberal. Given the excellent performance of the test based on SLRT-PB1 (in terms of power), confidence intervals based on the SLRT-PB1 statistic will also provide confidence intervals having the smallest expected width.

3. Likelihood based procedures for lognormal percentiles

We shall now consider inference concerning the pth percentile of the lognormal distribution. Thus the parameter of interest is $\eta = \mu + z_p\sigma$, where z_p denotes the pth percentile of the standard normal distribution. The log-likelihood function can be written as a function of η and σ^2, to be denoted as $l(\eta, \sigma^2)$:

$$l(\eta, \sigma^2) = \sum_{i=1}^{k} m_i \ln \Phi(z_{i\eta}) - (n-m)\ln(\sigma) - \frac{n-m}{2\sigma^2}\left(s_d^2 + (\bar{x}_d - \eta + z_p\sigma)^2\right), \tag{8}$$

where $z_{i\eta} = \frac{DL_i - \eta}{\sigma} + z_p$. The constrained MLE of σ^2, under the constraint $\eta = \eta_0$, can be obtained by maximizing the above likelihood function, after replacing η by η_0; for computational details, see the Appendix. The test statistics SLRT, SLRT-PB1 and SLRT-PB2 can now be computed following procedures similar to what is explained in the previous section.

Table 3 gives the type I error rates of the above three tests for testing the hypothesis $H_0 : \eta \geq \eta_0$ versus $H_0 : \eta < \eta_0$. We have used the same set up as in Table 1. Regarding the performance of the tests in terms of type I error probability, the conclusions are similar to those in the previous section. Table 4 gives the powers of three tests: the tests SLRT-PB1 and SLRT-PB2, as well as the test based on the generalized p-value, in the set up of Table 2. All three tests exhibit very similar performance in terms of power. In view of these conclusions regarding the performance of the three tests, the corresponding confidence intervals will exhibit similar performance in terms of coverage probability and expected width.

Remark 1. In Section 2 and Section 3, we have presented our results in the context of a Type I left-censored sample. If a sample is Type I right censored, then the procedure for a left-censored sample can be easily modified based on the symmetry of the normal distribution, since the log-transformed data multiplied by -1 results in a Type I left censored sample; see also Remark 1 in Krishnamoorthy, Mallick and Mathew (2011).

Table 1: Type I error rates of the SLRT, SLRT-PB1 (in parentheses) and SLRT-PB2 [in parenthesis] for testing $H_0 : \psi \geq \psi_0$ vs. $H_1 : \psi < \psi_0$; $\alpha = .05$; $\psi_0 = 3$

	$n = 10$		$n = 15$		$n = 20$	
p_1	σ=1	σ=3	σ=1	σ=3	σ=1	σ=3
.3	.068(.051)[.052]	.086(.049)[.052]	.067(.049)[.051]	.075(.049)[.047]	.065(.051)[.052]	.070(.047)[.051]
.5	.066(.051)[.049]	.090(.052)[.049]	.064(.048)[.046]	.074(.052)[.049]	.060(.050)[.046]	.072(.048)[.049]
.7	.000(.043)[.051]	.087(.050)[.051]	.026(.050)[.048]	.078(.048)[.050]	.036(.049)[.052]	.072(.052)[.050]
.8	.000(.000)[.047]	.090(.048)[.047]	.000(.053)[.047]	.079(.052)[.052]	.000(.044)[.045]	.074(.053)[.050]
(p_1, p_2)						
(.2,.4)	.075(.050)[.050]	.089(.052)[.051]	.063(.053)[.051]	.075(.048)[.051]	.064(.052)[.048]	.072(.049)[.048]
(.3,.6)	.071(.050)[.053]	.089(.053)[.047]	.064(.051)[.049]	.078(.052)[.049]	.062(.050)[.053]	.073(.050)[.053]
(.5,.6)	.064(.050)[.050]	.084(.051)[.046]	.064(.051)[.052]	.075(.053)[.053]	.058(.051)[.052]	.067(.047)[.050]
(.6,.8)	.026(.053)[.053]	.075(.051)[.049]	.039(.050)[.054]	.084(.047)[.050]	.041(.051)[.050]	.079(.051)[.050]
(p_1, p_2, p_3)						
(.1,.4,.5)	.077(.049)[.052]	.099(.052)[.053]	.072(.052)[.051]	.083(.049)[.051]	.064(.051)[.050]	.076(.051)[.050]
(.3,.5,.5)	.067(.051)[.049]	.090(.049)[.049]	.064(.049)[.049]	.076(.049)[.054]	.063(.048)[.050]	.073(.047)[.047]
(.4,.4,.7)	.064(.049)[.051]	.093(.050)[.045]	.065(.048)[.048]	.075(.052)[.051]	.062(.052)[.052]	.074(.052)[.052]
(.5,.6,.8)	.053(.051)[.047]	.088(.051)[.050]	.054(.050)[.046]	.081(.053)[.051]	.057(.050)[.047]	.074(.051)[.049]

Table 2: Powers of the GV test, SLRT-PB1 test (in parentheses) and SLRT-PB2 test [in parenthesis] for testing $H_0 : \psi \geq \psi_0$ vs. $H_a : \psi < \psi_0$ at 5% significance level

(a) $\mu = 0$ and $\psi_0 = 4.5$

	$n = 15$				$n = 30$			
(p_1, p_2, p_3)	σ=3	σ=2.5	σ=2	σ=1.5	σ=3	σ=2.5	σ=2	σ=1.5
(.2, .4, .5)	.049(.051)[.053]	.16(.17)[.15]	.42(.48)[.46]	.82(.89)[.86]	.049(.051)[.052]	.25(.29)[.26]	.69(.78)[.78]	.99(.99)[.99]
(.3, .4, .5)	.050(.051)[.051]	.15(.16)[.13]	.40(.42)[.36]	.77(.82)[.73]	.050(.050)[.053]	.25(.25)[.22]	.68(.72)[.65]	.98(.99)[.97]
(.3, .5, .6)	.051(.053)[.047]	.13(.15)[.13]	.36(.42)[.34]	.75(.79)[.69]	.047(.050)[.053]	.24(.24)[.20]	.63(.70)[.60]	.97(.98)[.95]
(.5, .7, .8)	.047(.047)[.046]	.13(.15)[.13]	.30(.36)[.30]	.67(.72)[.62]	.046(.049)[.049]	.19(.22)[.19]	.54(.62)[.53]	.89(.96)[.90]

(b) $\sigma = 1$ and $\psi_0 = 3.5$

	$n = 15$				$n = 30$			
(p_1, p_2, p_3)	μ=3	μ=2.5	μ=2	μ=1.5	μ=3	μ=2.5	μ=2	μ=1.5
(.2, .4, .5)	.049(.051)[.048]	.23(.31)[.30]	.42(.68)[.64]	.82(.87)[.86]	.049(.051)[.048]	.49(.56)[.54]	.92(.95)[.94]	.99(.99)[.99]
(.3, .4, .5)	.050(.051)[.053]	.22(.28)[.27]	.40(.63)[.54]	.77(.82)[.77]	.050(.050)[.051]	.48(.53)[.49]	.91(.92)[.89]	.98(.99)[.98]
(.3, .5, .6)	.051(.053)[.047]	.22(.28)[.27]	.36(.58)[.53]	.75(.79)[.72]	.047(.050)[.050]	.43(.51)[.49]	.87(.91)[.86]	.97(.99)[.97]
(.5, .7, .8)	.047(.047)[.049]	.19(.25)[.25]	.30(.54)[.49]	.67(.73)[.64]	.046(.049)[.050]	.37(.50)[.46]	.77(.87)[.84]	.91(.97)[.94]

Table 3: Type I error rates of the SLRT, SLRT-PB1 (in parentheses) and SLRT-PB2 [in parenthesis] for testing $H_0 : \eta \geq \eta_0$ vs. $H_a : \eta < \eta_0$; $\alpha = .05$; $\eta_0 = 3$

	$n = 10$		$n = 15$		$n = 20$	
p_1	σ=1	σ=3	σ=1	σ=3	σ=1	σ=3
.3	.081(.051)[.049]	.080(.050)[.048]	.069(.051)[.051]	.074(.051)[.046]	.068(.051)[.051]	.071(.053)[.050]
.5	.076(.049)[.049]	.079(.050)[.050]	.072(.052)[.050]	.074(.049)[.051]	.071(.048)[.051]	.073(.048)[.049]
.7	.069(.051)[.051]	.070(.052)[.047]	.072(.052)[.050]	.073(.052)[.048]	.073(.049)[.054]	.074(.052)[.051]
.8	.038(.047)[.053]	.040(.051)[.052]	.051(.046)[.049]	.050(.051)[.050]	.048(.054)[.048]	.054(.048)[.050]
(p_1, p_2)	σ=1	σ=3	σ=1	σ=3	σ=1	σ=3
(.2,.4)	.076(.049)[.049]	.082(.049)[.049]	.075(.049)[.050]	.076(.049)[.048]	.069(.048)[.049]	.072(.048)[.046]
(.3,.6)	.085(.048)[.051]	.082(.051)[.053]	.080(.051)[.050]	.071(.052)[.051]	.072(.053)[.050]	.067(.053)[.051]
(.5,.6)	.086(.052)[.052]	.082(.051)[.051]	.078(.051)[.050]	.073(.051)[.050]	.071(.050)[.048]	.072(.050)[.050]
(.4,.8)	.038(.051)[.049]	.075(.049)[.046]	.077(.048)[.053]	.074(.052)[.053]	.069(.049)[.048]	.071(.049)[.050]
(p_1, p_2, p_3)	$n = 15$		$n = 20$		$n = 30$	
(.2,.3,.4)	.074(.051)[.047]	.074(.051)[.049]	.076(.051)[.051]	.067(.051)[.048]	.068(.051)[.050]	.065(.051)[.052]
(.3,.5,.6)	.075(.047)[.047]	.081(.049)[.048]	.066(.049)[.051]	.069(.049)[.049]	.068(.048)[.045]	.069(.048)[.053]
(.5,.6,.7)	.078(.049)[.047]	.079(.052)[.048]	.072(.052)[.044]	.070(.052)[.046]	.066(.053)[.051]	.071(.053)[.048]
(.6,.7,.7)	.072(.051)[.054]	.074(.051)[.051]	.078(.051)[.047]	.077(.051)[.050]	.068(.050)[.051]	.070(.050)[.049]

Table 4: Powers of the GV test, the SLRT-PB1 test (in parentheses) and SLRT-PB2 [in parenthesis] for $H_0 : \eta \geq \eta_0$ vs. $H_a : \eta < \eta_0$ at 5% significance level when $\eta = \mu + z_{.9}\sigma$

(a) μ=1 and $\eta_0 = 4.846$

	$n = 15$				$n = 30$			
(p_1, p_2, p_3)	σ=3	σ=2.5	σ=2	σ=1.5	σ=3	σ=2.5	σ=2	σ=1.5
(.2, .4, .5)	.047(.050)[.050]	.12(.12)[.13]	.35(.36)[.36]	.80(.78)[.79]	.046(.050)[.051]	.22(.21)[.21]	.64(.64)[.65]	.99(.99)[.99]
(.3, .4, .5)	.052(.050)[.050]	.12(.12)[.13]	.34(.34)[.34]	.73(.76)[.78]	.051(.051)[.049]	.20(.20)[.21]	.59(.63)[.65]	.98(.98)[.98]
(.3, .5, .6)	.051(.047)[.051]	.12(.12)[.13]	.33(.34)[.34]	.72(.75)[.77]	.047(.048)[.049]	.20(.19)[.20]	.62(.63)[.61]	.97(.98)[.98]
(.5, .7, .8)	.047(.049)[.049]	.12(.12)[.12]	.32(.31)[.31]	.71(.69)[.68]	.047(.047)[.052]	.19(.19)[.20]	.57(.60)[.60]	.95(.97)[.97]

(b) σ=1 and $\eta_0 = 3.282$

	$n = 15$				$n = 30$			
(p_1, p_2, p_3)	μ=2	μ=1.5	μ=1	μ=.5	μ=2	μ=1.5	μ=1	μ=.5
(.2, .4, .5)	.047(.050)[.052]	.26(.26)[.26]	.62(.61)[.61]	.85(.88)[.87]	.049(.047)[.050]	.45(.47)[.47]	.92(.92)[.92]	.99(.99)[.99]
(.3, .4, .5)	.052(.050)[.049]	.25(.26)[.26]	.59(.61)[.61]	.83(.85)[.87]	.050(.049)[.047]	.46(.47)[.46]	.91(.91)[.91]	.99(.99)[.99]
(.3, .5, .6)	.051(.047)[.052]	.25(.26)[.26]	.57(.58)[.61]	.84(.84)[.85]	.047(.050)[.049]	.44(.44)[.45]	.91(.91)[.91]	.99(.99)[.99]
(.5, .7, .8)	.047(.049)[.049]	.23(.22)[.24]	.53(.52)[.50]	.80(.78)[.78]	.046(.053)[.046]	.41(.41)[.42]	.85(.86)[.88]	.99(.99)[.99]

4. Illustrative examples

We shall now present two examples in order to illustrate the results in the previous sections. Both the examples involve two detection limits each.

Example 1. We shall use the Atrazine concentration data as given in Table 9.7 of Helsel (2005, p. 159) to illustrate the procedures described in the preceding sections. The original data were altered by adding a second detection limit at .05 (see Helsel, 2005, p. 229). The probability plot in Figure 5.5 of Helsel (2005) indicates that lognormality assumption is tenable.

Table 5: Atrazine concentrations (μg/L) in a series of Nebraska wells before June

.38	< .05	< .01	.03	.03	.05	.02	< .01	< .01	< .01	.11	.09
< .01	< .01	< .01	< .01	.02	< .05	.02	.02	.05	.03	.05	< .01

For this data set, $n = 24$, $m_1 = 9$ and $m_2 = 2$. The MLEs based on the log-transformed data are $\widehat{\mu} = -4.206$, and $\widehat{\sigma} = 1.462$. Since the numbers of non-detects are $m_1 = 9$ and $m_2 = 2$, n_1 must be at least 9 and n_2 must be at least 2. We assume that $n_1 = 18$ and $n_2 = 6$. Using these values of n_1 and n_2, Krishnamoorthy and Xu's (2011) approximate pivotal method yielded a 95% upper confidence limit for $\mu - z_{.9}\sigma$ as .272. The SLRT-PB1 and SLRT-PB2 methods both yielded the same 95% upper confidence limit .268. The SLRT produced the value .236.

We also computed 95% confidence intervals for the mean Atrazine concentration. The SLRT-PB1 method produced a 95% confidence interval for the mean as (.021,.258), and the SLRT-PB2 method produced (.021, .256). The 95% upper confidence limits are .167 (SLRT-PB1) and .168 (SLRT-PB2). Krishnamoorthy and Xu (2011) have applied generalized variable approach to find confidence interval of the mean for this example. The reported 95% two-sided confidence interval is (.023, .247), and the 95% one-sided upper limit for the mean is .166.

Example 2. The data below are measurements of the concentration of zinc (micrograms per liter) in shallow ground water from Alluvial Fan zone in the San Joaquin Valley, California (Millard and Deverel, 1988). The original data were altered by ignoring one missing value and deleting an outlier 620. The normal probability plot of the resulting data, after log-transformation, indicated that a lognormal distribution is tenable. For this data, $n = 66$,

Table 6: Zinc concentrations ((μg/L) in groundwater at a geological zone in the San Joaquin valley, California

Alluvial Fan Zone								
<10	10	20	20	<10	11	10	33	17
9	10	20	10	10	19	<10	10	10
5	10	20	20	<10	8	10	20	<10
18	<10	<10	20	10	<3	10	10	10
<10	10	10	20	7	<10	20	10	20
12	<10	20	<10	<10	<10	20	10	29
10	10	40	20	10	<10	30	20	< 10
11	50	23						

and there are two detection limits on the original scale, namely 3 and 10, with $m_1 = 1$ and

$m_2 = 15$. The MLEs based on log-transformed data are $\widehat{\mu} = 2.456$ and $\widehat{\sigma} = .577$. We are interested in a 95% upper confidence limit for the 90th percentile of the zinc concentration distribution. The SLRT produced the upper limit as 29.22. The SLRT-PB1 yielded 29.55, and the SLRT-PB2 yielded 29.61. The generalized variable approach due to Krishnamoorthy and Xu (2011) produced 29.74.

5. Discussion

It is well known that accurate analysis of type I censored data is difficult, especially in the small sample scenario, even when a parametric assumption such as lognormality holds. As far as we are aware, satisfactory procedures are not available for computing confidence intervals and hypothesis tests, in spite of the extensive literature on the problem. Most of the literature deal with point estimation, but avoid discussion of confidence intervals and hypothesis tests. In an article commenting on the issue of non-detects, Helsel (2010, p. 261) notes that "Method evaluations for estimating a mean do not necessarily carry over to the more diffcult issues of how to compute interval estimates, upper percentiles, a correlation coeffecient, a regression slope and intercept, or a multidimensional surface when left censoring is present. There are many interesting issues still to be evaluated." In other words, Helsel is highlighting the lack of accurate confidence intervals and tests. Our work proposes the application of higher order modifications of the signed log-likelihood ratio test statistics for inference concerning the lognormal mean and percentiles, based on a type I censored sample resulting from the presence of multiple detection limits. The modifications we have used are given in DiCiccio, Martin and Stern (2001), and we could successfully apply them to the present investigation. The methodology we have developed is attractive from two perspectives: (i) they appear to be accurate regardless of the sample size, regardless of the number of detection limits, and regardless of the proportion of non-detects, and (ii) the required computation and implementation are simple and straightforward. Our overall conclusion is that the methodologies proposed in DiCiccio, Martin and Stern (2001) provide accurate inference concerning the lognormal mean and percentiles, and the procedures are easy to understand, as well as straightforward to apply. In short, our contribution demonstrates that an existing methodology can provide accurate analysis of data subject to non-detects, when no other accurate methods have been pointed out in the literature. Given the extensive applications of the lognormal distribution for the modeling and analysis of environmental as well as exposure data, and given the frequent occurrence of non-detects, it is hoped that this work will provide much needed framework and methodology for analyzing such data.

Appendix: Computation of MLE and constrained MLE

As the parametric bootstrap procedure requires repeated computations of the MLEs based on simulated samples, some details on computations of the MLEs and the constrained MLE are given below.

The MLEs $\widehat{\mu}$ and $\widehat{\sigma}$ are the roots of the equations,

$$f_1(\mu, \sigma) = \frac{\partial l(\mu, \sigma)}{\partial \mu} = (n - m)\frac{\bar{x}_d - \mu}{\sigma} - \sum_{i=1}^{k} m_i \xi(z_i^*) = 0$$

$$f_2(\mu,\sigma) = \frac{\partial l(\mu,\sigma)}{\partial \sigma} = -(n-m) + \frac{1}{\sigma^2}(n-m)(s_d^2 + (\bar{x}_d - \mu)^2) - \sum_{i=1}^{k} m_i z_i^* \xi(z_i^*) = 0,$$

where $z_i^* = (DL_i - \mu)/\sigma$ and $\xi(x) = \phi(x)/\Phi(x)$. These roots may be computed using the Newton-Raphson method. The necessary partial derivatives to implement the Newton-Raphson scheme are

$$\begin{aligned}
f_{1\mu}(\mu,\sigma) = \frac{\partial^2 l(\mu,\sigma)}{\partial \mu^2} &= -\frac{(n-m)}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{k} m_i F(z_i^*) \\
f_{1\sigma}(\mu,\sigma) = \frac{\partial^2 l(\mu,\sigma)}{\partial \mu \partial \sigma} &= -\frac{(n-m)(\bar{x}_d - \mu)}{\sigma^2} - \frac{1}{\sigma}\sum_{i=1}^{k} m_i z_i^* F(z_i^*) \\
f_{2\sigma}(\mu,\sigma) = \frac{\partial^2 l(\mu,\sigma)}{\partial \sigma^2} &= -\frac{2}{\sigma^3}(n-m)(s_d^2 + (\bar{x}_d - \mu)^2) + \frac{1}{\sigma}\sum_{i=1}^{k} m_i z_i^* \xi(z_i^*) \\
&- \frac{1}{\sigma}\sum_{i=1}^{k} m_i z_i^{*2} F(z_i^*),
\end{aligned}$$

where $F(x) = x\xi(x) + \xi^2(x)$. The Newton-Raphson iterative relation is given by

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \leftarrow \begin{pmatrix} \mu_0 \\ \sigma_0 \end{pmatrix} - \begin{pmatrix} f_{1\mu}(\mu_0,\sigma_0) & f_{1\sigma}(\mu_0,\sigma_0) \\ f_{1\sigma}(\mu_0,\sigma_0) & f_{2\sigma}(\mu_0,\sigma_0) \end{pmatrix}^{-1} \begin{pmatrix} f_1(\mu_0,\sigma_0) \\ f_2(\mu_0,\sigma_0) \end{pmatrix}, \tag{9}$$

where μ_0 and σ_0 are the initial guess values for the roots. The mean $\bar{x}_d$ and the standard deviation s_d based on the detected observations can be used as initial values for μ_0 and σ_0, respectively.

A.1 Calculation of the constrained MLE of σ^2 under the constraint $\psi = \mu + .5\sigma^2$

Letting $\lambda = \sigma^2$, it can be checked that the required partial differential equation $\frac{\partial l(\psi_0,\lambda)}{\partial \lambda} = h(\lambda) = 0$, where $l(\psi_0,\lambda)$ is given in (5), simplifies to

$$h(\lambda) = \frac{1}{2\lambda^{3/2}}\sum_{i=1}^{k} m_i \frac{\phi(z_{i\psi_0})}{\Phi(z_{i\psi_0})}\left(\frac{\lambda}{2} - (DL_i - \psi_0)\right) + \frac{(n-m)}{2\lambda^2}[(s_d^2 + (\bar{x}_d - \psi_0)^2) - \lambda^2/4 - \lambda] = 0.$$

Furthermore,

$$\begin{aligned}
h'(\lambda) &= \frac{1}{2\lambda^{3/2}}\sum_{i=1}^{k} m_i \frac{\phi(z_{i\psi_0})}{\Phi(z_{i\psi_0})}\left(\frac{3}{2}\left(\frac{DL_i - \psi_0}{\lambda}\right) - \frac{1}{4}\right) \\
&- \frac{1}{2\lambda^{3/2}}\sum_{i=1}^{k} m_i \left[\frac{\phi(z_{i\psi_0})}{\Phi(z_{i\psi_0})} z_{i\psi_0} + \left(\frac{\phi(z_{i\psi_0})}{\Phi(z_{i\psi_0})}\right)^2\right]\left(\frac{\lambda}{2} - (DL_i - \psi_0)\right) z'_{i\psi_0} \\
&- \frac{(n-m)}{\lambda^3}\left(s_d^2 + (\bar{x}_d - \psi_0)^2 - \frac{\lambda}{2}\right),
\end{aligned}$$

where $z'_{i\psi_0} = \frac{1}{2\lambda}\left(\frac{\sqrt{\lambda}}{2} - \frac{DL_i - \psi_0}{\sqrt{\lambda}}\right)$. To compute the constrained MLE $\sigma^2_{\psi_0}$, the following Newton-Raphson iterative scheme

$$\lambda \leftarrow \lambda - \frac{h(\lambda)}{h'(\lambda)},$$

with the variance s_d^2 as an initial value, can be used.

A.2 Computation of the constrained MLE of σ^2 under the constraint $\eta = \mu + z_p\sigma$

Let $\lambda = \sigma^2$. The partial derivative of (8) with respect to λ can be expressed as

$$h(\lambda) = \sqrt{\lambda}\sum_{i=1}^{k} m_i \frac{\phi(z_{i\eta})}{\Phi(z_{i\eta})}(DL_i - \eta) + (n-m)\left[\lambda - \sqrt{\lambda} z_p(\bar{x}_d - \eta) - s_d^2 - (\bar{x}_d - \eta)^2\right] = 0,$$

where $z_{i\eta} = \frac{DL_i - \eta}{\sqrt{\lambda}} + z_p$. The constrained MLE of λ is the root of the equation $h(\lambda) = 0$. To find this root using the Newton-Raphson method, the derivative

$$\begin{aligned} h'(\lambda) = \frac{\partial h(\lambda)}{\partial \lambda} &= \frac{1}{2\sqrt{\lambda}}\sum_{i=1}^{m} m_i(DL_i - \eta)\frac{\phi(z_{i\eta})}{\Phi(z_{i\eta})} + \frac{1}{2\lambda}\sum_{i=1}^{m} m_i(DL_i - \eta)^2\left[z_{i\eta}\frac{\phi(z_{i\eta})}{\Phi(z_{i\eta})} + \left(\frac{\phi(z_{i\eta})}{\Phi(z_{i\eta})}\right)^2\right] \\ &+ (n-m)\left[1 - \frac{1}{2\sqrt{\lambda}}(\bar{x}_d - \eta)\right]. \end{aligned}$$

The Newton-Raphson iterative scheme is

$$\lambda \leftarrow \lambda - \frac{h(\lambda)}{h'(\lambda)},$$

which can be implemented with the initial value s_d^2.

Calculation of $\xi(x) = \phi(x)/\Phi(x)$

We used "Intel Visual Fortran 11 with IMSL libraries" for all computations. During the iterative process, we noticed that the term $\xi(x) = \phi(x)/\Phi(x)$ may cause overflow error. This overflow error can be avoided as follows. Note that $\xi(x) \to 0$ as $x \to \infty$ and $\xi(x) \to \infty$ as $x \to -\infty$. The overflow errors occur for large negative values of x. This error can be avoided by defining $\xi(x) \simeq -x$ for $x < -20$ and $\xi(x) = \phi(x)/\Phi(x)$ for $-20 \le x < \infty$.

An alternative expression for $\xi(x)$, which somewhat speeds up the repeated computations of the MLEs and the constrained MLE, is based on the polynomial expression for the standard normal distribution function given in Hart et al. (1968, p. 137). Let

$P_0 = 913.167442114755700$, $P_1 = 1024.60809538333800$,
$P_2 = 580.109897562908800$, $P_3 = 202.102090717023000$,
$P_4 = 46.0649519338751400$, $P_5 = 6.81311678753268400$,
$P_6 = 6.047379926867041E-1$, $P_7 = 2.493381293151434E-2$,
and
$Q_0 = 1826.33488422951125$, $Q_1 = 3506.420597749092$,
$Q_2 = 3044.77121163622200$, $Q_3 = 1566.104625828454$,
$Q_4 = 523.596091947383490$, $Q_5 = 116.9795245776655$,
$Q_6 = 17.1406995062577800$, $Q_7 = 1.515843318555982$,
$Q_8 = 6.25E-2$.

Define

$$g(x) = \frac{P_7x^7 + P_6x^6 + P_5x^5 + P_4x^4 + P_3x^3 + P_2x^2 + P_1x + P_0}{Q_8x^8 + Q_7x^7 + Q_6x^6 + Q_5x^5 + Q_4x^4 + Q_3x^3 + Q_2x^2 + Q_1x + Q_0}$$

In terms of $g(x)$, the function $\xi(x)$ can be expressed as

$$\xi(x) = \begin{cases} \left(\frac{1}{\phi(x)} - \sqrt{2\pi}g(x)\right)^{-1}, & \text{for } \ x > 0, \\ \left(g(-x)\sqrt{2\pi}\ \right)^{-1}, & \text{for } \ x < 0. \end{cases}$$

References

Barndorff-Nielsen OE (1991). "Modified Signed Log-Likelihood Ratio." *Biometrika*, 78, 557-563.

Bhaumik DK, Kapur K, Bhaumik R, Reda DJ (2013). "Sample Size Determination and Hypothesis Testing for the Mean of a Lognormal Distribution." *Journal of Environmental Statistics*, **5**, 1-21.

DiCiccio TJ, Martin MA, Stern SE (2001). "Simple and Accurate One-Sided Inference From Signed Roots of Likelihood Ratios." *The Canadian Journal of Statistics*, **29**, 67-76.

Doganaksoy N, Schmee J (2000). "Practical Aspects of Correlated Likelihood Ratio Confidence Intervals: A Discussion of Jeng-Meeker and Wong-Wu." *Technometrics*, **42**, 156-159.

Fraser DAS, Reid N, Wu J (1999). "A Simple General Formula for Tail Probabilities for Frequentist and Bayesian Inference." *Biometrika*, **86**, 249-264.

Helsel DR (2005). *Nondetects and Data Analysis.* John Wiley, New York.

Helsel DR (2010). "Much Ado About Next to Nothing: Incorporating Nondetects in science." *Annals of Occupational Hygiene*, **54**, 257-262.

Jeng S-L, Meeker WQ (2000). "Comparisons of Approximate Confidence Interval Procedures for Type I Censored Data." *Technometrics*, **42**, 135-148.

Krishnamoorthy K, Mallick A, Mathew T (2011), "Inference for the Lognormal Mean and Quantiles Based on Samples With Left and Right Type I Censoring." *Technometrics*, **53**, 72-83.

Krishnamoorthy K, Mallick A, Mathew T (2009). "Model Based Imputation Approach for Data Analysis in the Presence of Non-Detectable Values." *Annals of Occupational Hygiene*, **59**, 249-268.

Krishnamoorthy K, 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.

Krishnamoorthy K, Mathew T (2009). *Statistical Tolerance Regions: Theory, Applications and Computation.* John Wiley, New York.

Krishnamoorthy K, Xu Z (2011). "Confidence Limits for Lognormal Percentiles and for Lognormal Mean Based on Samples With Multiple Detection Limits." *Annals of Occupational Hygiene*, **55**, 495-509.

Millard SP, Deverel SJ (1988). "Nonparametric Statistical Methods for Comparing Two Sites Based on Data With Multiple Nondetect Limits." *Water Resources Research*, **24**, 2087-2098.

Skovgaard IM (1996). "An Explicit Large Deviation Approximation to One Parameter Tests." *Bernoulli*, **2**, 145-165.

Wong ACM, Wu J (2000). "Practical Small Sample Asymptotics for Distributions Used in Life Data Analysis." *Technometrics*, **42**, 149-155.

Wu J, Wong ACM, Jiang G (2003). "Likelihood-based confidence intervals for a lognormal mean. *Statistics in Medicine*, **22**, 1849-1860.

Wu J, Wong ACM, Wei, W (2006). "Interval Estimation of the Mean Response in a Log-Regression Model." *Statistics in Medicine*, **25**, 2125-2135.

Affiliation:

Thomas Mathew
Department of Mathematics and Statistics
University of Maryland Baltimore County
Baltimore, MD 21250
E-mail: mathew@umbc.edu
URL: http://www.math.umbc.edu/~mathew

Journal of Environmental Statistics http://www.jenvstat.org
Volume 6, Issue 5 Submitted: 2013-09-27
August 2014 Accepted: 2014-07-22