Supplementary materials for this article are available online. Please click the Technometrics link at http://pubs.amstat.org.

Inference for the Lognormal Mean and Quantiles Based on Samples With Left and Right Type I Censoring

K. Krishnamoorthy and Avishek Mallick

Department of Mathematics
University of Louisiana at Lafayette
Lafayette, LA 70504
(*krishna@louisiana.edu*; *axm9357@louisiana.edu*)

Thomas Mathew

Department of Mathematics and Statistics
University of Maryland Baltimore County
Baltimore, MD 21250
(*mathew@umbc.edu*)

Interval estimation of the mean and quantiles of a lognormal distribution is addressed based on a Type I singly censored sample. A special case of interest is that of a sample containing values below a single detection limit. Generalized inferential procedures which use maximum likelihood estimation based approximate pivotal quantities, and some likelihood based methods, are proposed. The latter include methodology based on the signed log-likelihood ratio test (SLRT) statistic and the modified signed log-likelihood ratio test (MSLRT) statistic. The merits of the methods are evaluated for a left-censored sample using Monte Carlo simulation. For inference concerning the lognormal mean, the SLRT is to be preferred for left-tailed testing, generalized inference for right-tailed testing, and all three approaches provide nearly the same performance for two-tailed testing. These conclusions hold even when the proportion of censored values is as large as 0.70. For inference concerning quantiles, both the generalized inference approach and the MSLRT approach are satisfactory. In view of its simplicity and ease of understanding and implementation, the generalized inference procedure is to be preferred. The results are illustrated with two examples. Technical derivations are given on the *Technometrics* website as supplementary material.

KEY WORDS: Detection limit; Generalized inference; Generalized pivotal quantity; Modified signed likelihood ratio test; Signed likelihood ratio test; Tolerance interval.

1. INTRODUCTION

While analyzing environmental and exposure data, a very common phenomenon is the occurrence of nondetects, that is, observations below an analytical detection limit (*DL*), resulting in Type I singly left-censored samples. The presence of observations below the *DL* significantly complicates the data analysis. Faced with such data, several strategies have been recommended for data analysis. One approach consists of replacing the below *DL* values with a constant such as $DL/2$, and using methods available for complete samples. It is easy to demonstrate that the conclusions resulting from this routine practice can be seriously flawed; in fact, the conclusions may depend on the substitution value used for replacing sample values below the *DL*. In particular, Helsel (2005, table 7.1) has given an example where the confidence intervals come out drastically different depending on the substitution value used. Likelihood based methods and the imputation technique are two approaches that are more satisfactory; see Amemiya (1984), Lubin et al. (2004), Frome and Wambach (2005), and Krishnamoorthy, Mallick, and Mathew (2009). However, for obtaining likelihood based confidence intervals or test procedures, one has to rely on large sample theory, and the methodology may be unsatisfactory when the sample sizes are small and/or the percentage of nondetects is somewhat large. Modified likelihood methods have been proposed so as to obtain accurate inference when the sample sizes are not large. The modifications are based on recent research related to the likelihood, focusing on the development of small sample procedures; see Wong and Wu (2000) and Damilano and Puig (2002), where modified likelihood methods are developed in the context of Type I censored data. Another option is to use multiple imputation methods, that is, replace the nondetects with artificially generated data, so that complete sample methods can be used with suitable adjustments. In a recent article, Krishnamoorthy, Mallick, and Mathew (2009) have noted that the imputation approach can be calibrated to achieve a high degree of accuracy for the purpose of computing confidence intervals, tolerance intervals, prediction intervals, etc., for the normal and related distributions (such as the lognormal and gamma distributions). Yet another option is to use bootstrap methods. For estimating an exceedance probability, that is, the proportion of a population exceeding a threshold value, Wild et al. (1996) considered nonparametric bootstrap, parametric bootstrap, hybrid bootstrap, and Gibbs sampling methods for finding interval estimates when a left-censored sample is available from a lognormal distribution. The numerical studies in their article indicate that Gibbs sampling is the only satisfactory method for the specific problem that they have addressed. As of now, it is not clear if there is a single methodology that can be recommended to deal with the detection limit scenario for inference concerning the different parameters that are of interest in practical applications.

The present work is motivated by the need to have valid inference concerning two parameters that are frequently encountered in environmental and exposure data analysis: the arithmetic mean and the quantiles of a lognormal distribution. In the

TECHNOMETRICS, FEBRUARY 2011, VOL. 53, NO. 1
DOI 10.1198/TECH.2010.09189

absence of nondetects, satisfactory confidence intervals for the lognormal mean are derived in Krishnamoorthy and Mathew (2003) using the generalized confidence interval idea, and in Wu, Wong, and Jiang (2003) using modified likelihood methods; see also Wu, Wong, and Wei (2006). It appears that accurate inference on the lognormal mean is currently unavailable when nondetects are present, especially when the sample sizes are not large and the percentage of nondetects is somewhat large. Here we shall pursue the two approaches mentioned above: generalized variable (GV) methods, that is, methodology based on the concept of generalized confidence intervals, and modified likelihood methods. The implementation of the generalized variable method requires the construction of a *generalized pivotal quantity* (GPQ), and this is explained in the next section. As we shall see, we have actually constructed approximate GPQs, using approximate pivots based on the maximum likelihood estimators. The concept of a generalized confidence interval is due to Weerahandi (1993), and numerous applications are available in Weerahandi (1995, 2004). The procedure based on the modified signed likelihood ratio test (MSLRT) statistic that we have investigated is due to Fraser, Reid, and Wu (1999); see also Brazzale, Davison, and Reid (2007, chapter 8). The MSLRT statistic is obtained by suitably modifying the signed likelihood ratio test (SLRT) statistic. In fact, the MSLRT procedure that we have used for quantiles is already developed in Wong and Wu (2000). Here we would like to emphasize that even though the theory is quite generally available for the approaches we shall adopt in our work, the derivations and computations are specific to the problem at hand, and have to be carried out separately for each problem.

The article is organized as follows. In Section 2, we give a brief description of the computation of maximum likelihood estimates (MLEs), and some approximate pivotal quantities in the detection limit scenario, that is, for singly Type I left-censored data. We have then addressed inference for the lognormal mean when the sample is left censored. Section 3 contains the details concerning the generalized inference procedure as well as likelihood based procedures. Inference concerning quantiles is taken up in Section 4. For inference concerning the lognormal mean, numerical results are presented on the Type I error probabilities of tests, when the proportion of censored values (denoted by p_0) varies between 0.2 and 0.7. The Type I error probabilities are tabulated for one-sided testing as well as two-sided testing. Based on the numerical results, the conclusion is that the SLRT is to be preferred for left-tailed testing, the GV procedure is to be preferred for right-tailed testing, and all three approaches provide nearly the same performance for two-tailed testing. In our numerical results, we have also included the test based on the asymptotic normality of the MLE. However, this test turns out to be too liberal or too conservative (depending on the testing problem), and cannot be recommended for practical use. For inference concerning quantiles, there is very little difference between the GV approach and the MSLRT approach; both procedures are satisfactory. For one-sided testing for the lognormal mean, the unsatisfactory performance of the MSLRT, even compared to the SLRT, is somewhat surprising, since the MSLRT is expected to be more accurate. Clearly, the blind application of the MSLRT is to be approached with caution, since the technical conditions necessary for its validity are not entirely clear, especially when we have Type I censored data. Apart from its dependence on the model, the technical conditions may also depend on the particular inference problem; for example, the MSLRT is entirely satisfactory for inference concerning a lognormal quantile, but not so for the lognormal mean. The overall conclusion emerging from our work is that for a Type I singly left-censored sample, the SLRT is to be preferred for left-tailed testing for the lognormal mean. For right-tailed testing and two-tailed testing for the lognormal mean, and for inference concerning the quantiles, we recommend the GV approach. In spite of the satisfactory performance of the MSLRT in some cases, especially for quantiles, the GV approach is straightforward to understand and implement.

Based on a Type I censored sample from a normal distribution, a comparison of different procedures to compute confidence intervals for the mean and variance, as well as the quantiles, is investigated in Jeng and Meeker (2000); however, the lognormal mean is not addressed by them. The authors have made recommendations on the approach to be adopted for such interval estimation problems, and note that for the interval estimation of percentiles, the results are not entirely satisfactory when the data are Type I censored. We would like to emphasize that the results of Jeng and Meeker (2000) and Wong and Wu (2000), as well as our own results, are in the set up of a single censoring threshold; the situation of multiple censoring thresholds remains to be addressed. In their discussion of Jeng and Meeker (2000) and Wong and Wu (2000), Doganaksoy and Schmee (2000) emphasize "the ease of use and understanding by practitioners." The GV approach appears to have an edge in this regard.

2. PRELIMINARIES

Here we shall explain the computation of maximum likelihood estimates (MLEs), and the derivation of generalized pivotal quantities based on a Type I singly left-censored sample from a lognormal distribution. Clearly, it is enough to present the results in the context of a normal distribution $N(\mu, \sigma^2)$. Throughout, x_0 will denote a known censoring threshold.

2.1 Maximum Likelihood Estimation

For a sample of size n from $N(\mu, \sigma^2)$, let k denote the number of observations below the censoring threshold x_0, and let X_i $(i = 1, 2, \ldots, n-k)$ denote the observations above x_0. Cohen (1959, 1961) has derived the MLEs for the mean μ and variance σ^2, which can be computed numerically as solutions of some nonlinear equations. To present these equations, define

$$\xi = \frac{x_0 - \mu}{\sigma}, \qquad Z(\xi) = \frac{\phi(\xi)}{1 - \Phi(\xi)}, \qquad \text{and} \qquad Y(h, \xi) = \frac{hZ(-\xi)}{1-h}, \tag{1}$$

where ϕ and Φ denote, respectively, the density function and the distribution function of the standard normal distribution, and $h = k/n$. Thus h is the fraction of observations in the sample that is below x_0. Let

$$\bar{X}_l = \frac{1}{n-k}\sum_{i=1}^{n-k} X_i \qquad \text{and} \qquad S_l^2 = \frac{1}{n-k}\sum_{i=1}^{n-k}(X_i - \bar{X}_l)^2. \tag{2}$$

The MLEs of μ, σ^2, and ξ are the solutions of the equations

$$\begin{aligned}\mu &= \bar{X}_l - \varrho(h,\xi)(\bar{X}_l - x_0),\\ \sigma^2 &= S_l^2 + \varrho(h,\xi)(\bar{X}_l - x_0)^2,\\ \frac{1 - Y(h,\xi)(Y(h,\xi)-\xi)}{(Y(h,\xi)-\xi)^2} &= \frac{S_l^2}{(\bar{X}_l - x_0)^2},\end{aligned} \tag{3}$$

where $\varrho(h,\xi) = Y(h,\xi)/[Y(h,\xi)-\xi]$. Let $\widehat{\xi}$ be the solution of the third equation in (3). Then, the MLEs of μ and σ^2 can be computed by substituting $\varrho(h,\widehat{\xi})$ for $\varrho(h,\xi)$ in the first two equations of (3).

2.2 Approximate Pivotal Quantities

In order to motivate the approximate pivotal quantities that we shall use, let's first consider the case of a Type II censored sample. Recall that a Type II censored sample consists of observing k smallest order statistics in a sample of size n. Suppose we have a Type II censored sample from $\mathrm{N}(\mu,\sigma^2)$. An important result, useful for constructing confidence intervals and hypothesis tests, is that $\frac{\widehat{\mu}-\mu}{\sigma}$ and $\frac{\widehat{\sigma}}{\sigma}$ are pivotal quantities if the sample is Type II censored, where $\widehat{\mu}$ and $\widehat{\sigma}^2$ are the MLEs of μ and σ^2, respectively; see Lawless (2003, p. 562). This result implies that $\frac{\widehat{\mu}-\mu}{\widehat{\sigma}}$ is distributed as $\frac{\widehat{\mu}^*}{\widehat{\sigma}^*}$, and $\frac{\widehat{\sigma}}{\sigma}$ is distributed as $\widehat{\sigma}^*$, where $\widehat{\mu}^*$ and $\widehat{\sigma}^*$ are the MLEs based on a Type II censored sample from $\mathrm{N}(0,1)$ distribution. As a consequence, we can obtain the empirical distributions of the above pivotal quantities using Monte Carlo simulation. As an example, let c_1 and c_2 satisfy $P(c_1 \le \frac{\widehat{\mu}^*}{\widehat{\sigma}^*} \le c_2) = 1-\alpha$, where c_1 and c_2 can be obtained using Monte Carlo simulation. Then, $(\widehat{\mu} - c_2\widehat{\sigma}, \widehat{\mu} + c_1\widehat{\sigma})$ is a $100(1-\alpha)\%$ confidence interval for μ. Now if $\widehat{\mu}$ and $\widehat{\sigma}^2$ are the MLEs of μ and σ^2, respectively, based on a Type I censored sample from $\mathrm{N}(\mu,\sigma^2)$, then $\frac{\widehat{\mu}-\mu}{\sigma}$ and $\frac{\widehat{\sigma}}{\sigma}$ are not exact pivotal quantities, but they are approximately so, as noted in Schmee, Gladstein, and Nelson (1985). Thus, in the Type I censoring scenario, approximate confidence intervals for μ and σ^2 can be constructed based on the MLEs.

2.3 Approximate Generalized Pivotal Quantities (GPQs)

We shall first give a brief description of the generalized confidence interval idea, introduced by Weerahandi (1993). Let X be a random variable whose distribution depends on a scalar parameter of interest ψ and nuisance parameter δ, where δ could be a vector. Let x denote the observed value of X. A generalized pivotal quantity (GPQ) for ψ is a function of X, x, ψ, and δ, denoted by $G(X,x;\psi,\delta)$, and is required to satisfy the following conditions:

(i) given the observed value x, the distribution of $G(X,x;\psi,\delta)$ is free of any parameters,

(ii) the observed value of $G(X,x;\psi,\delta)$, namely $G(x,x;\psi,\delta)$, is free of the nuisance parameter δ.

When the above two conditions hold, let G_p denote the $100p$th percentile of $G(X,x;\psi,\delta)$. Then $G_{1-\alpha}$ can be used to obtain a $100(1-\alpha)\%$ one-sided upper confidence limit for ψ. A two-sided interval can be similarly defined. We note that while computing the above percentiles, x (the observed data) is to be treated as fixed. Numerous applications of generalized confidence intervals have appeared in the literature, especially for the interval estimation of parametric functions for which traditional solutions are difficult to obtain. Several such applications are given in the books by Weerahandi (1995, 2004). The asymptotic accuracy of a class of generalized confidence interval procedures has been established by Hannig, Iyer, and Patterson (2006).

In order to introduce the generalized confidence interval idea under Type I censoring, let $\widehat{\mu}_{\text{obs}}$ and $\widehat{\sigma}_{\text{obs}}$ be observed values of $\widehat{\mu}$ and $\widehat{\sigma}$, respectively, in a sample with nondetects. A GPQ for μ, say G_μ, will be a function of the random variables $(\widehat{\mu},\widehat{\sigma})$ and the observed values $(\widehat{\mu}_{\text{obs}},\widehat{\sigma}_{\text{obs}})$, and is required to satisfy the two conditions given above, where $X = (\widehat{\mu},\widehat{\sigma})$, $x = (\widehat{\mu}_{\text{obs}},\widehat{\sigma}_{\text{obs}})$, with μ taking the place of ψ, and the nuisance parameter δ being equal to σ.

Now consider G_μ given by

$$G_\mu = \widehat{\mu}_{\text{obs}} - \frac{\widehat{\mu}-\mu}{\widehat{\sigma}}\widehat{\sigma}_{\text{obs}} = \widehat{\mu}_{\text{obs}} - \frac{\widehat{\mu}^*}{\widehat{\sigma}^*}\widehat{\sigma}_{\text{obs}}, \tag{4}$$

where $\widehat{\mu}^*$ and $\widehat{\sigma}^*$ are defined in the preceding paragraph. It is easy to see that G_μ defined in (4) satisfies the two conditions for a GPQ stated above, approximately. It is the first condition that holds only approximately, since $\frac{\widehat{\mu}-\mu}{\widehat{\sigma}}$ is only an approximate pivot. In other words, G_μ is an approximate GPQ for μ. An approximate GPQ for σ can be obtained similarly, and is given by

$$G_\sigma = \frac{\sigma}{\widehat{\sigma}}\widehat{\sigma}_{\text{obs}} = \frac{\widehat{\sigma}_{\text{obs}}}{\widehat{\sigma}^*}. \tag{5}$$

The percentiles of G_μ and G_σ provide confidence intervals for μ and σ, respectively, referred to as generalized confidence intervals. We note that such generalized confidence intervals coincide with the approximate confidence intervals mentioned in Section 2.2. However, an observation that is useful to us is that a GPQ for a function $g(\mu,\sigma)$ is given by $g(G_\mu, G_\sigma)$. This result will be used in the sequel for developing approximate confidence limits for the mean and the quantiles of a lognormal distribution.

3. INFERENCE FOR A LOGNORMAL MEAN

3.1 Generalized Confidence Interval

The mean of a lognormal distribution is given by $\exp(\mu + \frac{\sigma^2}{2})$, where μ and σ^2 are the mean and variance, respectively, of the log-transformed random variable [which follows $\mathrm{N}(\mu,\sigma^2)$]. Let $\widehat{\mu}$ and $\widehat{\sigma}^2$ denote the MLEs of μ and σ^2, respectively, based on a singly Type I left-censored sample of size n from the lognormal distribution. Based on the arguments in Section 2.3, an approximate GPQ for $\psi = \mu + \frac{\sigma^2}{2}$ is given by

$$G_\psi = G_\mu + \frac{1}{2}(G_\sigma)^2 = \widehat{\mu}_{\text{obs}} - \frac{\widehat{\mu}^*}{\widehat{\sigma}^*}\widehat{\sigma}_{\text{obs}} + \frac{1}{2}\frac{\widehat{\sigma}^2_{\text{obs}}}{\widehat{\sigma}^{*2}}, \tag{6}$$

where $\widehat{\mu}_{\text{obs}}$ and $\widehat{\sigma}_{\text{obs}}$ are observed values of $\widehat{\mu}$ and $\widehat{\sigma}$, respectively. We note that $\widehat{\mu}_{\text{obs}}$ and $\widehat{\sigma}_{\text{obs}}$ can be computed as described in Section 2.1, where the data are now log-transformed. For a given $(\widehat{\mu}_{\text{obs}},\widehat{\sigma}_{\text{obs}})$, the following Monte Carlo method can be used to find confidence intervals for ψ.

Algorithm 1.

1. Find the number of values, say k, that are below the left-censoring threshold.
2. Compute the ML estimates $\widehat{\mu}_{\text{obs}}$ and $\widehat{\sigma}_{\text{obs}}$, as outlined in Section 2.1.
3. Generate a sample $Z_1, \dots, Z_n$ from a N(0, 1) distribution, and sort them in ascending order as $Z_{(1)}, \dots, Z_{(n)}$ so that $Z_{(1)}$ is the smallest.
4. Compute the MLEs $\widehat{\mu}^*$ and $\widehat{\sigma}^*$ based on $Z_{(k+1)}, \dots, Z_{(n)}$, and compute the GPQ G_ψ in (6).
5. Repeat steps 3 and 4 a large number of times, say, 10,000. Appropriate percentiles of G_ψ form a confidence interval for $\psi = \mu + \frac{\sigma^2}{2}$. More specifically, the lower and upper $\frac{\alpha}{2}$ quantiles of G_ψ form a $1-\alpha$ confidence interval for ψ.

Note that even though we are computing $\widehat{\mu}^*$ and $\widehat{\sigma}^*$ based on $Z_{(k+1)}, \dots, Z_{(n)}$, after discarding $Z_{(1)}, \dots, Z_{(k)}$, we are not in the set up of Type II censoring, since k is random and the value of k depends on the observed data.

3.2 Likelihood Based Approaches

Here we shall describe three likelihood based approaches: based on the asymptotic normality of the MLE, based on the normal approximation of the signed likelihood ratio test (SLRT) statistic, and based on the normal approximation of a modified signed likelihood ratio test (MSLRT) statistic. Before we provide the specifics on these as they apply to our problems, we shall give a brief description of the SLRT and MSLRT approaches. For details, we refer to the book by Brazzale, Davison, and Reid (2007).

Let $\theta = (\psi, \delta)$ be the parameter vector in an inference problem, and let $l(\psi, \delta)$ denote the log-likelihood function based on a sample $x_1, x_2, \dots, x_n$ of independent observations. Here ψ is a scalar parameter of interest and δ is a nuisance parameter. Let $(\widehat{\psi}, \widehat{\delta})$ denote the MLE of (ψ, δ), and let $\widehat{\delta}_\psi$ denote the constrained MLE of δ, for a fixed ψ. The signed likelihood ratio test is based on the result that

$$R(\psi) = \operatorname{sign}(\widehat{\psi} - \psi)\{2(l(\widehat{\psi}, \widehat{\delta}) - l(\psi, \widehat{\delta}_\psi))\}^{1/2} \sim \mathrm{N}(0, 1) \quad \text{asymptotically}, \tag{7}$$

where sign(y) is 1 if $y \ge 0$, and is -1 if $y < 0$. As is well known, tail area approximations based on the asymptotic normality of $R(\psi)$ are first order accurate, that is, the error is of the order $O(n^{-1/2})$.

Various modifications of the SLRT statistic are now available, with the goal of improving the accuracy of the normal approximation. The version that we shall use here is due to Fraser, Reid, and Wu (1999). In order to describe the modification due to these authors, let $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ denote the sample, and let $\mathbf{z} = (z_1(x_1, \theta), z_2(x_2, \theta), \dots, z_n(x_n, \theta))'$ be a vector of pivotal quantities, where $\theta = (\psi, \delta)$. We note that the pivot $z_i(x_i, \theta)$ is a function of only x_i (apart from the parameters). For example, one possible choice of $z_i(x_i, \theta)$ is simply the cumulative distribution function of x_i. Define

$$V = -\left(\frac{\partial \mathbf{z}}{\partial \mathbf{x}}\right)^{-1}\left(\frac{\partial \mathbf{z}}{\partial \theta^T}\right),$$

and let $V_1', \dots, V_n'$ denote the rows of V. Now define a *canonical parameter* $\varphi = \varphi(\theta)$ as

$$\varphi(\theta) = \sum_{i=1}^{n} \frac{\partial l(\theta; \mathbf{x})}{\partial x_i} V_i,$$

where $l(\theta; \mathbf{x})$ is the log-likelihood function, and $\varphi(\theta)$ also depends on $\mathbf{x}$. In order to introduce the modified statistic, let $\hat{\theta} = (\hat{\psi}, \hat{\delta})$ denote the MLE of θ, and write $\hat{\theta}_\psi = (\psi, \hat{\delta}_\psi)$. Also let $J_\theta(\theta)$ denote the observed information matrix and let $J_{\delta\delta}(\theta)$ denote the $(\delta\delta)$ block of $J_\theta(\theta)$. We shall also use the notation $\varphi_\theta(\theta)$ and $\varphi_\delta(\theta)$ to denote the derivatives of $\varphi(\theta)$ with respect to θ and δ, respectively. Now consider the statistic $R^*(\psi)$ defined as

$$R^*(\psi) = R(\psi) + \frac{1}{R(\psi)} \ln\left\{\frac{Q(\psi)}{R(\psi)}\right\}, \tag{8}$$

where

$$Q(\psi) = \frac{|\varphi(\hat{\theta}) - \varphi(\hat{\theta}_\psi),\ \varphi_\delta(\hat{\theta}_\psi)|}{|\varphi_\theta(\hat{\theta})|}\left\{\frac{|J_\theta(\hat{\theta})|}{|J_{\delta\delta}(\hat{\theta}_\psi)|}\right\}^{1/2}. \tag{9}$$

The tail area approximations based on the asymptotic normality of $R^*(\psi)$ are expected to be third-order accurate. In their work, Fraser, Reid, and Wu (1999) have rigorously investigated the improved accuracy of the normal approximation resulting from the use of $R^*(\psi)$.

We shall now return to the interval estimation of the lognormal mean based on the various approaches.

3.2.1 The MLE and Its Asymptotic Normality. For a singly Type I left-censored sample from the lognormal distribution, let x_0 denote the censoring threshold on the log scale, and let $\xi = (x_0 - \mu)/\sigma$. The log-likelihood function of the log-transformed observations is then given by

$$l(\mu, \lambda) = k \ln(\Phi(\xi)) - \frac{n-k}{2} \ln \lambda - \frac{(n-k)(S_l^2 + (\bar{X}_l - \mu)^2)}{2\lambda}, \tag{10}$$

where $\lambda = \sigma^2$ and $(\bar{X}_l, S_l^2)$ is as defined in (2), these quantities being defined in terms of the log-transformed data. To find the Fisher information matrix $J_{\mu,\lambda}$, we evaluated

$$-\frac{\partial^2 l(\mu, \lambda)}{\partial \mu^2} = \frac{k}{\lambda} w(\xi)[w(\xi) + \xi] + \frac{n-k}{\lambda},$$

$$-\frac{\partial^2 l(\mu, \lambda)}{\partial \mu\, \partial \lambda} = \frac{k w(\xi)}{2\lambda^{3/2}}\big[\xi(\xi + w(\xi)) - 1\big] + \frac{(n-k)(\bar{X}_l - \mu)}{\lambda^2},$$

$$-\frac{\partial^2 l(\mu, \lambda)}{\partial \lambda^2} = \frac{k\xi w(\xi)}{4\lambda^2}\big[\xi(\xi + w(\xi)) - 3\big] - \frac{n-k}{2\lambda^2} + \frac{(n-k)}{\lambda^3}(S_l^2 + (\bar{X}_l - \mu)^2),$$

where $w(\xi) = \phi(\xi)/\Phi(\xi)$. The observed information matrix

$$J_{\hat{\mu},\hat{\lambda}} = \begin{pmatrix} -\frac{\partial^2 l(\mu,\lambda)}{\partial \mu^2} & -\frac{\partial^2 l(\mu,\lambda)}{\partial \mu\, \partial \lambda} \\ -\frac{\partial^2 l(\mu,\lambda)}{\partial \mu\, \partial \lambda} & -\frac{\partial^2 l(\mu,\lambda)}{\partial \lambda^2} \end{pmatrix}\Bigg|_{\hat{\mu},\hat{\lambda}}.$$

Let $C = (c_{ij}) = J^{-1}_{\hat{\mu},\hat{\lambda}}$. Then

$$\frac{\hat{\mu} + \hat{\lambda}/2 - \psi}{\sqrt{c_{11} + c_{12} + c_{22}/4}} \sim \mathrm{N}(0,1) \quad \text{asymptotically}. \tag{11}$$

A confidence interval for ψ can be easily obtained based on the above asymptotic normal distribution.

3.2.2 The Signed Likelihood Ratio Test (SLRT) Statistic. Let $(\widehat{\psi}, \widehat{\sigma}^2)$ denote the MLE of (ψ, σ^2). Using (7), we note that the signed likelihood ratio test is based on the result that

$$\begin{aligned} R(\psi) &= \operatorname{sign}(\widehat{\psi} - \psi)\bigl\{2(l(\widehat{\psi}, \widehat{\sigma}^2) - l(\psi, \widehat{\sigma}^2_\psi))\bigr\}^{1/2} \\ &\sim \mathrm{N}(0,1) \quad \text{asymptotically}, \end{aligned} \tag{12}$$

where $\operatorname{sign}(y)$ is 1 if $y \geq 0$, and is -1 if $y < 0$ and $\widehat{\sigma}^2_\psi$ is the constrained MLE of σ^2 for a fixed ψ. In order to find the constrained MLE $\widehat{\sigma}^2_\psi$, we write the log-likelihood function (ignoring a constant) as

$$\begin{aligned} l(\psi, \sigma^2) &= k \ln \Phi(\xi_\sigma) - \frac{n-k}{2} \ln \sigma^2 \\ &\quad - \frac{(n-k)}{2\sigma^2}\left(S_l^2 + \left(\bar{X}_l - \psi + \frac{\sigma^2}{2}\right)^2\right), \end{aligned}$$

where $\xi^2_\sigma = (x_0 - \psi + \sigma^2/2)/\sigma$, with x_0 denoting the censoring threshold. The constrained MLE of $\widehat{\sigma}^2_\psi$ is the solution of the equation $\frac{\partial l(\psi,\sigma^2)}{\partial \sigma^2} = 0$, with respect to σ^2. Specifically, $\widehat{\sigma}^2_\psi$ is the solution to σ^2 of the equation

$$\begin{aligned} &-\frac{k}{2(\sigma^2)^{3/2}}(DL - \psi - \sigma^2/2)\frac{\phi(\xi^2_\sigma)}{\Phi(\xi^2_\sigma)} \\ &\quad + \frac{(n-k)}{2\sigma^4}\bigl(S_l^2 + (\bar{X}_l - \psi)^2 - \sigma^4/4 - \sigma^2\bigr) = 0. \end{aligned} \tag{13}$$

Approximate confidence limits for ψ can be obtained as the solutions to ψ, obtained by equating $R(\psi)$ to the appropriate standard normal percentiles.

3.2.3 Modified Signed Likelihood Ratio Test (MSLRT) Statistic. The MSLRT statistic $R^*(\psi)$ is defined in (8), where $R(\psi)$ is given in (12) and the factor $Q(\psi)$ is given below; a detailed derivation of $Q(\psi)$ is available online on the *Technometrics* website as supplemental material. Confidence intervals for ψ can be constructed using the asymptotic normality of $R^*(\psi)$. For the interval estimation of μ and the quantiles in the context of Type I censored data, the MSLRT approach is described in Wong and Wu (2000).

The observed information matrix $J_\theta(\theta)$ is necessary to obtain the expression for $Q(\psi)$. In order to give the expression for $J_\theta(\theta)$, let $\xi = (x_0 - \mu)/\sigma$, and $w(\xi) = \phi(\xi)/\Phi(\xi)$, x_0 being the left-censoring threshold. The expression for $J_\theta(\theta)$ is given by

$$J_\theta(\theta) = \begin{pmatrix} -\frac{\partial^2 l(\theta)}{\partial \mu^2} & -\frac{\partial^2 l(\theta)}{\partial \mu\, \partial \sigma} \\ -\frac{\partial^2 l(\theta)}{\partial \mu\, \partial \sigma} & -\frac{\partial^2 l(\theta)}{\partial \sigma^2} \end{pmatrix}, \tag{14}$$

where

$$\begin{aligned} -\frac{\partial^2 l(\mu,\sigma)}{\partial \mu^2} &= \frac{k}{\sigma^2} w(\xi)[w(\xi) + \xi] + \frac{n-k}{\sigma^2}, \\ -\frac{\partial^2 l(\mu,\sigma)}{\partial \mu\, \partial \sigma} &= \frac{k w(\xi)}{\sigma^2}\bigl[\xi(\xi + w(\xi)) - 1\bigr] \\ &\quad + \frac{2(n-k)(\bar{X}_l - \mu)}{\sigma^3}, \\ -\frac{\partial^2 l(\mu,\sigma)}{\partial \sigma^2} &= \frac{k\xi w(\xi)}{\sigma^2}\bigl[\xi(\xi + w(\xi)) - 2\bigr] - \frac{n-k}{\sigma^2} \\ &\quad + \frac{3(n-k)}{\sigma^4}(S_l^2 + (\bar{X}_l - \mu)^2), \end{aligned} \tag{15}$$

k being the number of observations below the censoring threshold x_0, and $\bar{X}_l$ and S_l^2 being the quantities defined in (2).

The computation of $Q(\psi)$ for the lognormal mean can now be accomplished by following the steps given below:

1. For $\theta = (\mu, \sigma^2)$, let $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma}^2)$ be the maximum likelihood estimator of θ.
2. Keeping $\psi = \mu + \frac{\sigma^2}{2}$ fixed, let $\widehat{\sigma}^2_\psi$ denote the maximum likelihood estimator of σ^2, obtained by solving (13). With $\widehat{\sigma}^2_\psi$ so obtained, let $\widehat{\mu}_\psi = \psi - \widehat{\sigma}^2_\psi/2$, and $\widehat{\theta}_\psi = (\widehat{\mu}_\psi, \widehat{\sigma}^2_\psi)$.
3. Let
$$\hat{\alpha} = \frac{k}{\widehat{\sigma}_\psi}\frac{\phi((x_0 - \widehat{\mu}_\psi)/\widehat{\sigma}_\psi)}{\Phi((x_0 - \widehat{\mu}_\psi)/\widehat{\sigma}_\psi)} - \frac{(n-k)(\bar{X}_l - \widehat{\mu}_\psi)}{\widehat{\sigma}^2_\psi},$$
where ϕ and Φ denote the standard normal density and distribution function, respectively, and $\bar{X}_l$ is the sample mean defined in (2).
4. With the 2×2 information matrix $J_\theta(\theta)$ defined in Equations (14) and (15), compute $J_\theta(\widehat{\theta})$ and $J_\theta(\widehat{\theta}_\psi)$, which are the values of $J_\theta(\theta)$ evaluated at $\widehat{\theta}$ and $\widehat{\theta}_\psi$, respectively. Let $J_{\theta,ij}(\widehat{\theta}_\psi)$ denote the (ij)th element of $J_\theta(\widehat{\theta}_\psi)$, for i, $j = 1, 2$.
5. In terms of the quantities defined above, the factor $Q(\psi)$ is given by
$$\begin{aligned} Q(\psi) &= \operatorname{sign}(\widehat{\psi} - \psi)\left|\frac{\widehat{\sigma}^2_\psi}{\widehat{\sigma}^2}\left(\widehat{\mu} - \widehat{\mu}_\psi - \frac{\widehat{\sigma}^2_\psi}{2}\right) + \frac{\widehat{\sigma}^2_\psi}{2}\right| \\ &\quad \times |J_\theta(\widehat{\theta})|^{1/2}(\widehat{\sigma}/\widehat{\sigma}_\psi)^5 \\ &\quad /\{J_{\theta,22}(\widehat{\theta}_\psi) - \widehat{\alpha} - 2J_{\theta,12}(\widehat{\theta}_\psi)\widehat{\sigma}_\psi \\ &\quad + J_{\theta,11}(\widehat{\theta}_\psi)\widehat{\sigma}^2_\psi\}^{1/2}. \end{aligned} \tag{16}$$

3.3 Simulation Studies

In order to assess the accuracy of the inference procedures for the lognormal mean, proposed in the previous sections, we simulated the Type I error probability of tests for $\psi = \mu + \frac{1}{2}\sigma^2$ based on the following four procedures: (i) test based on the generalized variable (GV) method, that is, based on the generalized confidence interval in Section 3.1, (ii) the test (denoted by AN) based on the asymptotic normality of the MLE of the lognormal mean, (iii) the SLRT based on the asymptotic normality of $R(\psi)$ given in (12), and (iv) the MSLRT based on the asymptotic normality of $R^*(\psi)$ given in (8), where the factor $Q(\psi)$ is defined in (16). The simulations were carried out after specifying p_0, the proportion of observations below the censoring threshold x_0, for $p_0 = 0.2, 0.3, 0.5$, and 0.7, and for the parameter values $\mu = 0$ and $\sigma = 1, 2, 3$. The Type I error rates were simulated for one-tailed as well as two-tailed tests. The test based on the GV approach rejects the null hypothesis when the boundary point of the null is not included in the relevant one-sided or two-sided generalized confidence interval. Since the other three tests are based on normal approximations, the rejection rule consists of evaluating the test statistics, and comparing them with the appropriate standard normal percentiles. While carrying out the simulations, we have taken the boundary of the null hypothesis to be $\mu + \frac{1}{2}\sigma^2$ for $\mu = 0$ and for $\sigma = 1, 2, 3$. We used Fortran programs for all computations, and they can be obtained from the first author upon request.

Type I error rates of the tests are reported in Table 1 for the parameter combinations mentioned above for sample size $n = 20, 30$, and 50. The performance of the tests depends on the values of σ and p_0, and also on the type of test: left-tailed, right-tailed, or two-tailed. No test emerges as the clear winner in all the scenarios. The test based on the asymptotic normality of the MLE (denoted by AN in Table 1) seems to be the worst among all tests. This test exhibits poor performance even for large samples. For example, when $n = 100$, $p_0 = 0.50$, and $\sigma = 2$, the Type I error rates are 0.012, 0.103, and 0.072 for left-tailed, right-tailed, and two-tailed tests, respectively; they are 0.004, 0.114, and 0.083 when $p_0 = 0.70$. Thus, the test AN should be avoided in applications.

For left-tailed testing, the SLRT appears to exhibit the most satisfactory performance, even though it is conservative in some cases. Both the GV approach and the MSLRT appear to be quite liberal in most cases, especially when σ is large. For right-tailed testing, it is the GV approach that provides the most satisfactory solution. For two-tailed testing, all the three procedures provide satisfactory performance, even though the SLRT is liberal and the MSLRT is conservative in a few cases. The GV approach appears to be quite satisfactory for two-sided testing, for all the cases considered in the simulation. As the sample size gets large, there is very little difference between the GV approach and the MSLRT. In several instances, the MSLRT is not as satisfactory as the SLRT. This is somewhat surprising since the normal approximation for the MSLRT statistic is expected to be more accurate than for the SLRT. This may be due to the violation of certain technical conditions required for the validity of the asymptotics for the MSLRT; however, the theoretical underpinnings are not clear to us.

Since the tests produce very different Type I error probabilities, power comparisons are not meaningful in all the scenarios reported in Table 1. However, we have carried out a limited power study for a few cases where the Type I error probabilities are almost equal. The powers are reported in Table 2 for the tests GV, SLRT and MSLRT for two-sided testing of the lognormal mean for the parameter value $\sigma^2 = 1$, and null value $\psi_0 = 0.5$. Thus the entries in Table 2 corresponding to $\psi = 0.5$ are the Type I error probabilities (for a 5% nominal level). From Table 2, we see that the powers of the three tests are somewhat similar, and no test emerges as a clear choice.

4. INFERENCE FOR QUANTILES

We shall now discuss the computation of confidence limits for the quantiles of a lognormal distribution in the scenario of Type I left censoring. Clearly, it is enough to construct confidence limits for the quantiles of the corresponding normal distribution. Let z_p denote the pth quantile of the standard normal distribution. For the normal distribution $\mathrm{N}(\mu, \sigma^2)$, we shall investigate the computation of a $100(1-\alpha)\%$ upper confidence limit for $\mu + z_p\sigma$, and a $100(1-\alpha)\%$ lower confidence limit for $\mu - z_p\sigma$. The upper confidence limit so constructed is also referred to as an upper tolerance limit for $\mathrm{N}(\mu, \sigma^2)$. We recall that an upper tolerance limit for $\mathrm{N}(\mu, \sigma^2)$ is such that at least a proportion p of $\mathrm{N}(\mu, \sigma^2)$ is below the limit, with confidence $1-\alpha$. Similarly, a lower tolerance limit for $\mathrm{N}(\mu, \sigma^2)$ is such that at least a proportion p of $\mathrm{N}(\mu, \sigma^2)$ is above the limit, with confidence $1-\alpha$. The computation of such a lower tolerance limit reduces to the computation of a $100(1-\alpha)\%$ lower confidence limit for $\mu - z_p\sigma$. In other words, the computation of confidence limits for the quantiles is important since they provide tolerance limits. We refer to Krishnamoorthy and Mathew (2009) for further details on tolerance limits.

We shall follow the notations in Section 2 and Section 3. Let $\eta = \mu + z_p\sigma$, and we shall discuss the computation of an upper confidence limit for η using the generalized variable (GV) approach, and using the SLRT statistic and the MSLRT statistic.

4.1 Generalized Variable Approach

Similar to (6), we get the following approximate GPQ for η:

$$\begin{aligned} G_\eta &= \widehat{\mu}_{\mathrm{obs}} - \frac{\widehat{\mu} - \mu}{\widehat{\sigma}}\widehat{\sigma}_{\mathrm{obs}} + z_p\frac{\sigma}{\widehat{\sigma}}\widehat{\sigma}_{\mathrm{obs}} \\ &= \widehat{\mu}_{\mathrm{obs}} - \frac{\widehat{\mu}^*}{\widehat{\sigma}^*}\widehat{\sigma}_{\mathrm{obs}} + z_p\frac{\widehat{\sigma}_{\mathrm{obs}}}{\widehat{\sigma}^*} \\ &= \widehat{\mu}_{\mathrm{obs}} + \frac{z_p - \widehat{\mu}^*}{\widehat{\sigma}^*}\widehat{\sigma}_{\mathrm{obs}}, \end{aligned} \tag{17}$$

where the various quantities in (17) are as defined in Sections 2 and 3. Let $w_{p;\alpha}$ denote the αth quantile of $w_p = \frac{z_p - \widehat{\mu}^*}{\widehat{\sigma}^*}$. Then

$$\widehat{\mu}_{\mathrm{obs}} + w_{p;1-\alpha}\widehat{\sigma}_{\mathrm{obs}} \tag{18}$$

is a $100(1-\alpha)\%$ upper confidence limit for $\mu + z_p\sigma$. Proceeding as above, it can be verified that

$$\widehat{\mu}_{\mathrm{obs}} + w_{1-p;\alpha}\widehat{\sigma}_{\mathrm{obs}} \tag{19}$$

is a $100(1-\alpha)\%$ lower confidence limit for $\mu - z_p\sigma$.

Algorithm 2. Based on a Type I left-censored sample, the following steps can be used to find a $100(1-\alpha)\%$ upper confidence limit for $\mu + z_p\sigma$.

1. Let k be the number of left-censored observations.
2. Compute the ML estimates $\widehat{\mu}_{\mathrm{obs}}$ and $\widehat{\sigma}_{\mathrm{obs}}$.
3. Generate a sample $Z_1, \ldots, Z_n$ from a N(0, 1) distribution, and sort them in ascending order as $Z_{(1)}, \ldots, Z_{(n)}$ so that $Z_{(1)}$ is the smallest.
4. Compute the MLEs $\widehat{\mu}^*$ and $\widehat{\sigma}^*$ based on $Z_{(k+1)}, \ldots, Z_{(n)}$, and compute $w_p = \frac{z_p - \widehat{\mu}^*}{\widehat{\sigma}^*}$.
5. Repeat steps 3–4 a large number of times, say, 10,000.

The $100(1-\alpha)$ percentile of the 10,000 w_p's so generated is an estimate of $w_{p;1-\alpha}$ in (18), and the upper confidence limit is given by $\widehat{\mu}_{\mathrm{obs}} + w_{p;1-\alpha}\widehat{\sigma}_{\mathrm{obs}}$.

Table 1. Type I error rates of the tests for a lognormal mean under Type I censoring for the choice $\mu = 0$; L = left-tailed test, R = right-tailed test, T = two-tailed test (p_0 is the proportion of left-censored observations)

		$\sigma = 1$			$\sigma = 2$			$\sigma = 3$		
p_0	Method	L	R	T	L	R	T	L	R	T
						$n = 20$				
0.2	GV	0.054	0.044	0.050	0.060	0.050	0.052	0.049	0.049	0.053
	AN	0.021	0.105	0.082	0.004	0.124	0.091	0.002	0.152	0.115
	SLRT	0.046	0.069	0.059	0.043	0.040	0.040	0.035	0.077	0.062
	MSLRT	0.053	0.042	0.047	0.057	0.039	0.049	0.062	0.038	0.052
0.3	GV	0.053	0.044	0.058	0.059	0.040	0.052	0.055	0.049	0.050
	AN	0.021	0.110	0.083	0.004	0.137	0.102	0.001	0.151	0.117
	SLRT	0.044	0.058	0.051	0.034	0.064	0.051	0.040	0.063	0.051
	MSLRT	0.053	0.036	0.043	0.060	0.037	0.046	0.064	0.038	0.052
0.5	GV	0.059	0.036	0.044	0.080	0.033	0.053	0.055	0.054	0.055
	AN	0.022	0.108	0.085	0.002	0.157	0.122	0.000	0.176	0.139
	SLRT	0.048	0.059	0.056	0.041	0.066	0.058	0.039	0.070	0.056
	MSLRT	0.049	0.029	0.036	0.060	0.028	0.051	0.078	0.027	0.053
0.7	GV	0.062	0.016	0.043	0.070	0.023	0.053	0.051	0.044	0.046
	AN	0.039	0.003	0.017	0.001	0.197	0.164	0.000	0.213	0.184
	SLRT	0.052	0.040	0.043	0.043	0.070	0.056	0.033	0.073	0.054
	MSLRT	0.044	0.030	0.037	0.079	0.014	0.046	0.102	0.015	0.061
						$n = 30$				
0.2	GV	0.054	0.047	0.054	0.055	0.046	0.049	0.060	0.044	0.052
	AN	0.028	0.096	0.071	0.009	0.119	0.084	0.006	0.130	0.091
	SLRT	0.046	0.060	0.055	0.043	0.065	0.056	0.037	0.066	0.052
	MSLRT	0.051	0.044	0.047	0.056	0.041	0.049	0.055	0.042	0.049
0.3	GV	0.056	0.046	0.054	0.060	0.045	0.050	0.054	0.040	0.048
	AN	0.025	0.092	0.069	0.009	0.126	0.092	0.003	0.138	0.104
	SLRT	0.050	0.061	0.054	0.041	0.064	0.052	0.039	0.067	0.053
	MSLRT	0.055	0.036	0.044	0.063	0.041	0.052	0.065	0.039	0.050
0.5	GV	0.060	0.039	0.052	0.062	0.033	0.050	0.061	0.033	0.043
	AN	0.024	0.098	0.076	0.004	0.141	0.107	0.002	0.149	0.116
	SLRT	0.052	0.039	0.051	0.047	0.066	0.057	0.036	0.068	0.054
	MSLRT	0.048	0.033	0.039	0.064	0.034	0.047	0.073	0.033	0.052
0.7	GV	0.066	0.024	0.048	0.073	0.040	0.053	0.075	0.022	0.047
	AN	0.039	0.042	0.023	0.002	0.168	0.138	0.000	0.190	0.156
	SLRT	0.045	0.040	0.042	0.040	0.062	0.053	0.036	0.069	0.052
	MSLRT	0.046	0.027	0.032	0.073	0.018	0.046	0.088	0.018	0.055
						$n = 50$				
0.2	GV	0.053	0.053	0.048	0.051	0.050	0.049	0.055	0.041	0.044
	AN	0.028	0.080	0.060	0.017	0.100	0.067	0.012	0.108	0.073
	SLRT	0.043	0.059	0.055	0.043	0.060	0.054	0.043	0.063	0.056
	MSLRT	0.051	0.047	0.049	0.058	0.044	0.050	0.057	0.048	0.053
0.3	GV	0.067	0.044	0.049	0.066	0.047	0.063	0.054	0.053	0.057
	AN	0.026	0.083	0.060	0.014	0.101	0.071	0.008	0.115	0.081
	SLRT	0.049	0.059	0.058	0.041	0.057	0.051	0.044	0.059	0.054
	MSLRT	0.050	0.047	0.048	0.056	0.041	0.048	0.057	0.039	0.049
0.5	GV	0.061	0.041	0.052	0.066	0.032	0.051	0.062	0.044	0.049
	AN	0.027	0.086	0.065	0.008	0.125	0.095	0.004	0.124	0.095
	SLRT	0.049	0.058	0.051	0.044	0.064	0.055	0.042	0.067	0.056
	MSLRT	0.050	0.039	0.043	0.062	0.036	0.052	0.062	0.037	0.051
0.7	GV	0.079	0.015	0.048	0.080	0.040	0.054	0.073	0.022	0.053
	AN	0.035	0.056	0.040	0.004	0.143	0.110	0.001	0.151	0.119
	SLRT	0.048	0.045	0.047	0.042	0.058	0.049	0.037	0.066	0.052
	MSLRT	0.047	0.041	0.042	0.065	0.024	0.044	0.071	0.025	0.049

Table 2. Powers of the two-tailed tests for a lognormal mean under Type I left censoring; $H_0: \psi = \psi_0$ vs. $H_a: \psi \neq \psi_0$; $\psi_0 = 0.5$, $\sigma^2 = 1$, $\alpha = 0.05$ (p_0 is the proportion of left-censored observations)

		$n=30$							$n=40$						
		ψ							ψ						
p_0	Method	0.5	0.6	0.7	0.8	0.9	1.0	1.2	0.5	0.6	0.7	0.8	0.9	1.0	1.2
0	GV	0.05	0.10	0.18	0.36	0.60	0.90	0.99	0.05	0.11	0.24	0.45	0.72	0.91	0.99
	SLRT	0.05	0.08	0.18	0.35	0.59	0.91	0.99	0.05	0.09	0.22	0.46	0.76	0.98	1
	MSLRT	0.05	0.09	0.19	0.35	0.59	0.91	0.99	0.05	0.20	0.26	0.44	0.76	0.93	0.99
0.2	GV	0.05	0.10	0.18	0.36	0.57	0.82	0.96	0.05	0.11	0.25	0.44	0.71	0.92	0.98
	SLRT	0.06	0.08	0.18	0.33	0.55	0.83	0.95	0.05	0.09	0.21	0.45	0.76	0.97	1
	MSLRT	0.05	0.10	0.18	0.36	0.55	0.82	0.97	0.05	0.20	0.25	0.44	0.69	0.90	0.99
0.3	GV	0.05	0.10	0.18	0.33	0.56	0.82	0.93	0.05	0.10	0.24	0.43	0.71	0.89	0.97
	SLRT	0.05	0.08	0.17	0.33	0.55	0.82	0.92	0.05	0.08	0.21	0.43	0.71	0.95	1
	MSLRT	0.05	0.10	0.18	0.33	0.54	0.79	0.90	0.05	0.18	0.24	0.42	0.67	0.87	0.99
0.5	GV	0.05	0.09	0.17	0.32	0.54	0.70	0.92	0.05	0.09	0.24	0.39	0.64	0.80	0.97
	SLRT	0.06	0.08	0.17	0.32	0.50	0.70	0.87	0.06	0.08	0.20	0.39	0.62	0.83	0.99
	MSLRT	0.05	0.09	0.17	0.32	0.52	0.70	0.89	0.05	0.17	0.24	0.39	0.60	0.79	0.97
0.7	GV	0.05	0.09	0.17	0.31	0.47	0.62	0.87	0.05	0.09	0.23	0.38	0.53	0.69	0.94
	SLRT	0.05	0.08	0.16	0.31	0.48	0.61	0.81	0.05	0.08	0.19	0.38	0.56	0.71	0.88
	MSLRT	0.05	0.09	0.17	0.31	0.45	0.62	0.86	0.05	0.15	0.23	0.36	0.52	0.67	0.83

4.2 Likelihood Approaches

Methodology based on the SLRT and MSLRT statistics is presented in Wong and Wu (2000), and we refer to this article for details. Here we shall only outline the steps necessary to implement them. More details are available online as supplemental material, on the *Technometrics* website.

4.2.1 Signed Likelihood Ratio Test (SLRT) Statistic. The signed likelihood ratio test is based on the result that

$$R(\eta) = \mathrm{sign}(\widehat{\eta} - \eta)\left\{2(l(\widehat{\eta}, \widehat{\sigma}^2) - l(\eta, \widehat{\sigma}_\eta^2))\right\}^{1/2} \sim \mathrm{N}(0, 1) \quad \text{asymptotically}, \tag{20}$$

where $\widehat{\sigma}_\eta^2$ is the constrained MLE of σ^2 when η is kept fixed. To find this constrained MLE, we write the log-likelihood function in terms of η and σ^2 as

$$l(\eta, \sigma^2) = k \ln \Phi(\xi_\sigma^*) - \frac{n-k}{2} \ln \sigma^2 - \frac{(n-k)}{2\sigma^2}\left(S_l^2 + (\bar{X}_l - \eta + z_p\sigma)^2\right),$$

where $\xi_\sigma^* = (x_0 - \eta)/\sigma + z_p$, x_0 being the censoring threshold. The constrained MLE of σ^2 is the solution of the equation

$$\frac{\partial l(\eta, \sigma^2)}{\partial \sigma^2} = -k\sigma(x_0 - \eta)w(\xi_\sigma^*) + (n-k) \times [S_l^2 + (\bar{X}_l - \eta)^2 + z_p\sqrt{\sigma^2}(\bar{X}_l - \eta) - \sigma^2] = 0, \tag{21}$$

where $w(\xi_\sigma^*) = \phi(\xi_\sigma^*)/\Phi(\xi_\sigma^*)$, $\phi(\cdot)$ and $\Phi(\cdot)$ denoting the density function and the cdf, respectively, of the standard normal distribution.

4.2.2 Modified Signed Likelihood Ratio Test (MSLRT) Statistic. With $\theta = (\mu, \sigma)$, we note that the parameter of interest is $\eta = \eta(\theta) = \mu + z_p\sigma$. Let $\hat{\theta} = (\widehat{\mu}, \widehat{\sigma})$ denote the MLE of θ. Then the MLE of η is $\widehat{\eta} = \widehat{\mu} + z_p\widehat{\sigma}$. Let $\widehat{\sigma}_\eta^2$ denote the constrained MLE of σ^2, as defined above. Then the constrained MLE of θ, say $\widehat{\theta}_\eta$, is given by

$$\widehat{\theta}_\eta = (\eta - z_p\widehat{\sigma}_\eta, \widehat{\sigma}_\eta) = (\widehat{\mu}_\eta, \widehat{\sigma}_\eta).$$

The constrained MLE can also be obtained by maximizing $l(\eta, \sigma^2) + \alpha(\eta(\theta) - \eta_0)$, where α is a Lagrange multiplier and η_0 is a fixed value of $\eta = \eta(\theta)$.

In order to give the expression for the MSLRT statistic, let $J_\theta(\theta)$ be as defined in Equations (14) and (15). Let

$$\widehat{\sigma}_\chi^2 = (1, z_p)J_\theta^{-1}(\widehat{\theta}_\eta)\begin{pmatrix} 1 \\ z_p \end{pmatrix}, \tag{22}$$

and define

$$Q(\eta) = \mathrm{sign}(\widehat{\eta} - \eta)\left|\frac{\widehat{\sigma}_\eta^2}{\widehat{\sigma}^2}\left(\widehat{\mu} - \widehat{\mu}_\eta - \frac{\widehat{\sigma}_\eta z_p}{2}\right) + \frac{\widehat{\sigma}_\eta z_p}{2}\right| \times \left\{\frac{1}{\widehat{\sigma}_\chi^2} \times \left(\frac{\widehat{\sigma}^2}{\widehat{\sigma}_\eta^2}\right)^5\right\}^{1/2}. \tag{23}$$

The MSLRT statistic is then given by

$$R^*(\eta) = R(\eta) + \frac{1}{R(\eta)} \ln \frac{Q(\eta)}{R(\eta)}.$$

More details on the derivation of $Q(\eta)$ are available online as supplemental material, on the *Technometrics* website.

Table 3. Coverage probabilities of 95% lower confidence limits for $\mu - z_{0.9}\sigma$; $\mu = 0$ (p_0 is the proportion of left-censored observations)

		$n = 20$			$n = 30$			$n = 50$		
		σ			σ			σ		
p_0	Method	1	2	3	1	2	3	1	2	3
0.2	GV	0.953	0.951	0.952	0.952	0.951	0.952	0.951	0.953	0.951
	SLRT	0.961	0.960	0.965	0.959	0.960	0.961	0.960	0.958	0.959
	MSLRT	0.951	0.956	0.956	0.953	0.952	0.951	0.952	0.951	0.948
0.3	GV	0.955	0.956	0.955	0.954	0.954	0.954	0.950	0.952	0.951
	SLRT	0.966	0.965	0.966	0.960	0.962	0.962	0.959	0.963	0.961
	MSLRT	0.955	0.956	0.954	0.957	0.954	0.954	0.953	0.949	0.953
0.5	GV	0.955	0.956	0.958	0.956	0.957	0.957	0.955	0.957	0.957
	SLRT	0.968	0.973	0.968	0.962	0.965	0.966	0.961	0.963	0.961
	MSLRT	0.956	0.958	0.960	0.958	0.959	0.957	0.956	0.956	0.956
0.7	GV	0.968	0.968	0.970	0.961	0.960	0.959	0.960	0.960	0.956
	SLRT	0.974	0.973	0.973	0.972	0.965	0.969	0.960	0.965	0.967
	MSLRT	0.960	0.960	0.961	0.960	0.957	0.956	0.960	0.963	0.960

4.3 Coverage Studies

Table 3 gives the coverage probabilities for a 95% lower confidence limit for $\mu - z_{0.9}\sigma$, and Table 4 gives the coverage probabilities for a 95% upper confidence limit for $\mu + z_{0.9}\sigma$, for the same sample size, parameter combinations, and p_0 values considered in Table 1. The coverage probability results in Tables 3 and 4 are for the intervals computed using the generalized variable (GV) approach, using the SLRT approach, and using the MSLRT approach. The SLRT based interval is conservative for the lower confidence limit, and liberal for the upper confidence limit. There is very little difference in coverage between the intervals obtained using the GV approach and using the MSLRT approach. Both of them are quite accurate in meeting the coverage probability requirement. From the description of the GV approach, it should be clear that the procedure is simple to understand, and easy to implement.

Table 5 gives the powers of the tests carried out using the GV approach, and using the SLRT and MSLRT approaches for testing hypothesis concerning the 90th percentile of the lognormal distribution, so that $\eta = \mu + 1.281\sigma$. A right-tailed test is considered with null value equal to 1.281 for the parameter choice $\sigma^2 = 1$. From the numerical results, it is clear that in terms of power, both the SLRT and MSLRT have a clear edge. The MSLRT gives more power for alternatives close to the null value, and the SLRT dominates in terms of power for alternatives farther away from the null value.

Remark 1. We have presented our results in the context of a singly Type I left-censored sample. If a sample is singly Type I right censored, then the procedure for a left-censored sample can be easily modified based on the symmetry of the normal distribution. For example, in order to obtain a 95% upper confidence limit for the 90th percentile, namely, $\mu + z_{0.90}\sigma$, we first compute a 95% lower confidence limit for the 10th percentile based on the negative log-transformed data, and multiply the result by -1 to get the desired upper confidence limit.

Remark 2. In general, the SLRT and the MSLRT approaches are applicable to any parametric distribution, where the interval

Table 4. Coverage probabilities of 95% upper confidence limits for $\mu + z_{0.9}\sigma$; $\mu = 0$ (p_0 is the proportion of left-censored observations)

		$n = 20$			$n = 30$			$n = 50$		
		σ			σ			σ		
p_0	Method	1	2	3	1	2	3	1	2	3
0.2	GV	0.941	0.942	0.942	0.943	0.944	0.943	0.945	0.945	0.945
	SLRT	0.931	0.930	0.930	0.934	0.931	0.931	0.938	0.936	0.936
	MSLRT	0.948	0.947	0.943	0.948	0.947	0.947	0.951	0.945	0.948
0.3	GV	0.942	0.942	0.943	0.945	0.945	0.945	0.946	0.946	0.946
	SLRT	0.927	0.928	0.928	0.938	0.934	0.934	0.934	0.939	0.939
	MSLRT	0.949	0.951	0.942	0.947	0.946	0.945	0.949	0.945	0.950
0.5	GV	0.947	0.947	0.948	0.947	0.946	0.947	0.949	0.949	0.949
	SLRT	0.929	0.927	0.927	0.933	0.935	0.935	0.940	0.935	0.935
	MSLRT	0.944	0.944	0.947	0.949	0.941	0.947	0.944	0.948	0.944
0.7	GV	0.951	0.952	0.952	0.950	0.950	0.950	0.951	0.950	0.951
	SLRT	0.926	0.921	0.921	0.923	0.927	0.927	0.935	0.928	0.928
	MSLRT	0.954	0.955	0.958	0.952	0.948	0.954	0.952	0.952	0.948

Table 5. Powers of the right-tailed tests for the 90th percentile; $H_0: \eta = \eta_0$ vs. $H_a: \eta > \eta_0$; $\eta_0 = 1.281$, $\delta = |\eta - \eta_0|$, $\sigma^2 = 1$, $\alpha = 0.05$ (p_0 is the proportion of left-censored observations)

		$n=30$							$n=40$						
		δ							δ						
p_{DL}	Method	0	0.1	0.2	0.3	0.4	0.5	0.6	0	0.1	0.15	0.2	0.3	0.4	0.5
0	GV	0.05	0.08	0.15	0.28	0.46	0.64	0.80	0.05	0.09	0.18	0.35	0.56	0.75	0.89
	SLRT	0.06	0.08	0.16	0.32	0.63	0.90	0.97	0.06	0.08	0.19	0.45	0.86	0.97	0.99
	MSLRT	0.05	0.20	0.22	0.34	0.57	0.82	0.95	0.05	0.20	0.26	0.42	0.72	0.93	0.97
0.2	GV	0.05	0.08	0.15	0.28	0.46	0.64	0.80	0.04	0.08	0.18	0.35	0.55	0.75	0.89
	SLRT	0.04	0.10	0.15	0.32	0.60	0.90	0.96	0.04	0.12	0.19	0.28	0.60	0.93	0.98
	MSLRT	0.05	0.19	0.22	0.33	0.54	0.80	0.94	0.05	0.19	0.25	0.41	0.71	0.92	0.96
0.3	GV	0.05	0.08	0.15	0.27	0.45	0.64	0.78	0.05	0.08	0.18	0.35	0.55	0.74	0.88
	SLRT	0.04	0.08	0.15	0.30	0.57	0.86	0.94	0.04	0.11	0.18	0.27	0.58	0.91	0.98
	MSLRT	0.05	0.19	0.22	0.32	0.53	0.75	0.88	0.05	0.19	0.24	0.41	0.68	0.92	0.92
0.5	GV	0.04	0.07	0.14	0.27	0.44	0.63	0.76	0.04	0.07	0.17	0.33	0.54	0.73	0.87
	SLRT	0.04	0.09	0.14	0.28	0.49	0.75	0.87	0.04	0.11	0.17	0.26	0.50	0.81	0.95
	MSLRT	0.05	0.18	0.21	0.31	0.49	0.71	0.81	0.05	0.18	0.24	0.39	0.62	0.85	0.88
0.7	GV	0.03	0.06	0.12	0.21	0.37	0.52	0.66	0.04	0.07	0.16	0.32	0.47	0.63	0.76
	SLRT	0.05	0.08	0.13	0.25	0.41	0.66	0.72	0.04	0.10	0.16	0.23	0.43	0.71	0.81
	MSLRT	0.05	0.17	0.20	0.29	0.40	0.55	0.70	0.05	0.17	0.23	0.36	0.52	0.69	0.80

estimation of a real-valued function of parameters is of interest. However, the GV approach that we have developed here depends on the observation that $\frac{\widehat{\mu}-\mu}{\sigma}$ and $\frac{\widehat{\sigma}}{\sigma}$ are approximate pivotal quantities under the scenario of Type I censoring; these being exact pivots if the sample is Type II censored. This observation concerning a Type II censored sample is valid for any distribution belonging to the location-scale family, where μ is the location parameter and σ is the scale parameter; see Lawless (2003, appendix E). Recently, Krishnamoorthy and Xie (2011) have noted that the approximate pivots produce satisfactory tolerance limits for double exponential and logistic distributions, which are location-scale distributions. Thus we expect that $\frac{\widehat{\mu}-\mu}{\sigma}$ and $\frac{\widehat{\sigma}}{\sigma}$ are approximate pivots when detection limits are present in samples from any location-scale family of distributions. However, we have not rigorously investigated this.

5. EXAMPLES

Here we shall present two examples. Example 1 is a direct illustration of our results for analyzing lognormal data in the presence of a single detection limit, so that we have a singly Type I left-censored sample. Since our results are also applicable for analyzing right-censored data (as noted in Remark 1), Example 2 deals with such a scenario. The necessary computations for the examples were carried out using Fortran programs, and they can be obtained from the first author upon request.

5.1 Example 1

This example is taken from Wild et al. (1996), and the data represent oil mist measurements obtained from a machining workshop in France. Oil mist is an aerosol of machining fluids suspected to induce several respiratory diseases. As noted in Wild et al. (1996), for the purpose of enforcing safety regulations in the context of occupational health, a parameter of interest is the exceedance probability, that is, the probability of oil mist levels exceeding a specified threshold limit. Inference concerning this parameter actually reduces to inference concerning a percentile of the underlying distribution; see Krishnamoorthy and Mathew (2009, section 1.1.3). Thus inference concerning a percentile is quite relevant for health hazard evaluation.

The 14 oil mist measurements (uncensored) are

1.7, 1.8, 2.1, 2.3, 2.3, 2.5, 2.8, 2.9, 2.9, 3.0, 3.0, 3.8, 3.8, 5.3.

Based on a normal probability plot of the log-transformed data, we conclude that the sample is from a lognormal distribution. The mean and standard deviation of the log-transformed data are $\bar{x} = 1.0097$ and $s = 0.3060$.

To illustrate the methods in the detection limit situation, let us assume a detection limit of ln(2.4) for the log-transformed data, so that ln(2.4) is the left-censoring threshold. Using the log-transformed data that are above ln(2.4), the MLEs are computed as $\widehat{\mu} = 0.9928$ and $\widehat{\sigma} = 0.3208$. In order to obtain a 95% upper confidence limit for $\mu + z_{0.90}\sigma$, we proceeded as follows. Starting with $\widehat{\mu} + z_{0.90}\widehat{\sigma}$, a forward search for η_0 was done until the p-value for testing $H_0: \eta \geq \eta_0$ versus $H_a: \eta < \eta_0$ is equal 0.05. This method using SLRT yielded an upper confidence limit for $\mu + z_{0.90}\sigma$ as 1.674, and hence the upper confidence limit on the original scale is computed as $\exp(1.674) = 5.333$. Using the MSLRT, we obtained the upper confidence limit as $\exp(1.777) = 5.912$. In order to apply the generalized confidence interval approach, we computed the 95th percentile of $(z_p - \widehat{\mu}^*)/\widehat{\sigma}^*$ (using simulation with 10,000 runs) as 2.410. The required upper confidence limit is obtained as $\exp(\widehat{\mu} + 2.410\widehat{\sigma}) = \exp(0.9928 + 2.410 \times 0.3208) = \exp(1.766) = 5.848$. We note that the upper confidence limit resulting from the MSLRT, and that resulting from the generalized confidence interval methodology are in close agreement, as noted earlier based on the numerical results in Table 3.

To judge the loss of information due to censoring, and to compare the results of censored and uncensored samples, we

Table 6. 95% confidence limits for the lognormal mean

Approach	Lower limit	Upper limit	Two-sided interval
GV	2.387	3.466	(2.291, 3.662)
SLRT	2.416	3.350	(2.326, 3.494)
MSLRT	2.361	3.432	(2.261, 3.618)

shall compute the upper confidence limit for $\mu + z_{0.9}\sigma$ based on all sample observations.

An Upper Confidence Limit for $\mu + z_{0.9}\sigma$ Based on the Complete Sample. If there is no detection limit, the 95% upper confidence limit is $\bar{x} + k_1 \times s$, where the factor k_1 is given by

$$\frac{1}{\sqrt{14}} t_{13;0.95}(z_{0.9}\sqrt{14}) = 2.1088.$$

Here $t_{m;0.95}(\delta)$ denotes the 95th percentile of a noncentral t distribution with df = m and noncentrality parameter δ. Thus the 95% upper confidence limit for $\mu + z_{0.9}\sigma$ is

$$\bar{x} + 2.1088 \times s = 1.0097 + 2.1088 \times 0.3060 = 1.655.$$

Consequently the upper confidence limit in the original scale is $\exp(1.655) = 5.233$. We note that, as expected, the upper confidence limit based on the complete sample is smaller than that based on the sample with a detection limit.

Confidence Intervals for the Mean. One-sided and two-sided confidence limits for the mean, that is, for $\exp(\mu + \frac{\sigma^2}{2})$, have been computed based on the three approaches, for a nominal level of 95%, where the generalized confidence interval was obtained using 10,000 simulation runs. The results are in Table 6.

The intervals obtained using the SLRT are narrower since the interval is liberal, as noted based on our numerical results. We also note that the confidence limits obtained using the MSLRT, and those obtained using the generalized variable approach are in good agreement.

5.2 Example 2

The data in Table 7 represent failure mileages (in units of 1000 miles) of different locomotive controls in a life test involving 96 locomotive controls. The test was terminated after 135,000 miles, and by then 37 controls had failed. Thus, we have a Type I right-censored sample of 37 failure mileages. Schmee and Nelson (1977) and Lawless (2003, section 5.3) have used the data to illustrate lognormal based inferential methods. In order to assess the reliability of the controls, it is of interest to find a lower confidence limit for the 10th percentile $\exp(\mu - z_{0.90}\sigma)$. This will provide some idea on how low the failure mileages can go.

Table 7. Failure mileages (in 1000) of locomotive controls

22.5	37.5	46.0	48.5	51.5	53.0	54.5	57.5	66.5	68.0
69.5	76.5	77.0	78.5	80.0	81.5	82.0	83.0	84.0	91.5
93.5	102.5	107.0	108.5	112.5	113.5	116.0	117.0	118.5	119.0
120.0	122.5	123.0	127.5	131.0	132.5	134.0			

As already noted, the sample is Type I right censored with censoring mileage of 135,000. Following the argument in Remark 1, MLEs based on the log-transformed data were computed as $\widehat{\mu}_0 = 5.117$ and $\widehat{\sigma}_0 = 0.705$. The percentile $w_{0.90;0.95}$ appearing in (19) is evaluated (with $n = 96$ and $l = 59$) as 1.579. Thus, using (19), we get $5.117 - 1.579 \times 0.705 = 4.004$, and $\exp(4.004) = 54.82$ is the desired 95% lower confidence limit for $\exp(\mu - z_{0.90}\sigma)$. Thus with 95% confidence we can assert that at least 90% of the controls survive 54,820 miles.

In order to implement the SLRT and MSLRT approaches, we proceeded similar to Example 1. The 95% lower confidence limit for $\mu - z_{0.90}\sigma$ came out to be 4.021 and 3.997, based on the SLRT statistic and the MSLRT statistic, respectively. Thus, on the original scale, the corresponding lower confidence limits are 55.57 and 54.435. We once again notice the agreement between the generalized confidence limit and that obtained using the MSLRT statistic.

6. CONCLUDING REMARKS

In general, Type I censoring presents difficulties for the data analysis, even for distributions such as the normal and the lognormal. Data analysis under Type I censoring is investigated in this article, for inference concerning the arithmetic mean and the quantiles of a lognormal distribution. The purpose of our investigation was to explore the applicability of some of the novel methodologies that are currently available in the literature, namely, procedures based on the generalized variables (GV) approach, and based on the modified signed likelihood ratio test (MSLRT) statistic. Published literature indicates that both of these approaches have found fruitful applications in a number of inference problems, especially when small sample inference is desired. In our investigation we have compared the performance of these two procedures along with that of the signed likelihood ratio test (SLRT) statistic for inference concerning the mean and quantiles of a lognormal distribution. Based on numerical results, specific recommendations are made on the procedure to be used for practical applications. In particular, the generalized confidence interval and the MSLRT based confidence interval are both satisfactory for inference concerning a lognormal quantile, when the percentage of censored observations can be as large as 70%. However, the conclusion is not so clear cut for inference on the lognormal mean. In fact, our work shows that the routine application of the MSLRT must be approached with caution; the procedure may even give results that are less satisfactory than the solution based on the SLRT. A final point to note is that the GV approach is easier to understand and implement, especially for a practitioner, and it provides accuracy very similar to that of the MSLRT for estimating the lognormal quantiles. For right-tailed testing or two-tailed testing for the lognormal mean, the GV approach was once again found satisfactory. Thus for such applications, we recommend the GV methodology in the context of the lognormal samples under Type I censoring. We also note that the GV methodology that we have developed for the lognormal distribution is applicable to any location scale family since approximate pivots can be developed in a similar fashion.

In view of the work that we have carried out, a problem that suggests itself is that of data analysis in the presence of multiple censoring thresholds, for example, multiple detection limits.

An MLE based methodology can obviously be implemented in a straightforward manner, as done in Section 3.2.1. However, given the poor performance of the resulting procedure even in the case of a single threshold, methodology based on the other approaches needs to be developed, and its performance has to be numerically studied. In fact, the various methods that currently exist to handle singly Type I censored data, do not address the case of multiple censoring limits. This is currently under investigation.

SUPPLEMENTARY MATERIALS

Appendix: An appendix providing a detailed derivation of the MSLRT statistics for the lognormal mean and qantiles, under a Type I singly left-censored sample, is available on the *Technometrics* web site. (Appendix.pdf)

ACKNOWLEDGEMENTS

This work has benefitted greatly from the comments and suggestions by two reviewers, an Associate Editor and the Editor. We are grateful to them for their input. This research work was supported by National Institute of Occupational Safety and Health grant R01-OH03628.

[Received November 2009. Revised October 2010.]

REFERENCES

Amemiya, T. (1984), "Tobit Models: A Survey," *Journal of Econometrics*, 24, 3–61. [72]

Brazzale, A. R., Davison, A. C., and Reid, N. (2007), *Applied Asymptotics: Case Studies in Small-Sample Statistics*, Cambridge: Cambridge University Press. [73,75]

Cohen, A. C. (1959), "Simplified Estimators for the Normal Distribution When Samples Are Singly Censored or Truncated," *Technometrics*, 1, 217–237. [73]

——— (1961), "Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples," *Technometrics*, 3, 535–541. [73]

Damilano, G., and Puig, P. (2002), "Small Sample Asymptotics for Type I Censored Data," *Biometrical Journal*, 44, 867–876. [72]

Doganaksoy, N., and Schmee, J. (2000), "Practical Aspects of Corrected Likelihood Ratio Confidence Intervals: A Discussion of Jeng–Meeker and Wong–Wu," *Technometrics*, 42, 156–159. [73]

Fraser, D. A. S., Reid, N., and Wu, J. (1999), "A Simple General Formula for Tail Probabilities for Frequentist and Bayesian Inference," *Biometrika*, 86, 249–264. [73,75]

Frome, E. L., and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data With Non-Detectable Values," preprint, U.S. Department of Energy. Available at *www.hss.energy.gov/nhealthsafety/ iipp/hservices/statmethods.pdf*. [72]

Hannig, J., Iyer, H. K., and Patterson, P. (2006), "Fiducial Generalized Confidence Intervals," *Journal of the American Statistical Association*, 101, 254–269. [74]

Helsel, D. R. (2005), *Nondetects and Data Analysis*, New York: Wiley. [72]

Jeng, S.-L., and Meeker, W. Q. (2000), "Comparisons of Approximate Confidence Interval Procedures for Type I Censored Data," *Technometrics*, 42, 135–148. [73]

Krishnamoorthy, K., and Mathew, T. (2003), "Inferences on the Means of Lognormal Distributions Using Generalized p-Values and Generalized Confidence Intervals," *Journal of Statistical Planning and Inference*, 115, 103–121. [73]

——— (2009), *Statistical Tolerance Regions: Theory, Applications and Computation*, New York: Wiley. [77,81]

Krishnamoorthy, K., and Xie, F. (2011), "Tolerance Intervals for Symmetric Location-Scale Families Based on Uncensored or Censored Samples," *Journal of Statistical Planning and Inference*, 141, 1170–1182. [81]

Krishnamoorthy, K., Mallick, A., and Mathew, T. (2009), "Model Based Imputation Approach for Data Analysis in the Presence of Non-Detectable Values: Normal and Related Distributions," *Annals of Occupational Hygiene*, 59, 249–268. [72]

Lawless, J. F. (2003), *Statistical Models and Methods for Lifetime Data*, New York: Wiley. [74,81,82]

Lubin, J. H., Colt, J. S., Camann, D., Davis, S., Cerhan, J. R., Severson, R. K., Bernstein, L., and Hartge, P. (2004), "Epidemiologic Evaluation of Measurement Data in the Presence of Detection Limits," *Environmental Health Perspectives*, 112, 1691–1696. [72]

Schmee, J., and Nelson, W. (1977), "Estimates and Approximate Confidence Limits for (Log) Normal Life Distributions From Singly Censored Samples by Maximum Likelihood," TIS Report 76CRD250, General Electric Co., Corporate Research and Development. [82]

Schmee, J., Gladstein, D., and Nelson, W. (1985), "Confidence Limits of a Normal Distribution From Singly Censored Samples Using Maximum Likelihood," *Technometrics*, 27, 119–128. [74]

Weerahandi, S. (1993), "Generalized Confidence Intervals," *Journal of the American Statistical Association*, 88, 899–905. [73]

——— (1995), *Exact Statistical Methods for Data Analysis*, New York: Springer-Verlag. [73,74]

——— (2004), *Generalized Inference in Repeated Measures: Exact Methods in MANOVA and Mixed Models*, New York: Wiley. [73,74]

Wild, P., Hordan, R., Leplay, A., and Vincent, R. (1996), "Confidence Intervals for Probabilities of Exceeding Threshold Limits With Censored Lognormal Data," *Environmetrics* 17, 247–259. [72,81]

Wong, A. C. M., and Wu, J. (2000), "Practical Small Sample Asymptotics for Distributions Used in Life Data Analysis," *Technometrics*, 42, 149–155. [72,73,76,79]

Wu, J., Wong, A. C. M., and Jiang, G. (2003), "Likelihood-Based Confidence Intervals for a Lognormal Mean," *Statistics in Medicine*, 22, 1849–1860. [73]

Wu, J., Wong, A. C. M., and Wei, W. (2006), "Interval Estimation of the Mean Response in a Log-Regression Model," *Statistics in Medicine*, 25, 2125–2135. [73]

Supplemental Material

Inference for the Lognormal Mean and Quantiles Based on Samples with Left and Right Type I Censoring

K. Krishnamoorthy and **Avishek Mallick**
Department of Mathematics
University of Louisiana at Lafayette
Lafayette, LA 70504, USA

Thomas Mathew
Department of Mathematics and Statistics
University of Maryland Baltimore County
Baltimore, MD 21250, USA

APPENDIX: DERIVATION OF THE MSLRT STATISTICS

In this Appendix, we provide a detailed derivation of the MSLRT statistics for the lognormal mean and qauntiles, under a type I singly left censored sample.

A1. The MSLRT statistic for the lognormal mean

The mean of a lognormal distribution is given by $\exp(\mu + \frac{\sigma^2}{2})$, where μ and σ^2 are the mean and variance, respectively, of the log-transformed random variable (which follows $N(\mu, \sigma^2)$). Define $\psi = \mu + \frac{\sigma^2}{2}$ and consider the reparametrization $(\mu, \sigma^2) \rightarrow (\psi, \sigma^2)$. Let $(\widehat{\psi}, \widehat{\sigma}^2)$ denote the MLE of (ψ, σ^2), and let $\widehat{\sigma}^2_\psi$ denote the constrained MLE of σ^2 for a fixed ψ. The computation of the MLE is described in Section 2.1, and that of the constrained MLE is explained in Section 3.2.2 in the paper. In particular, $\widehat{\sigma}^2_\psi$ is the solution of equation (13) in the paper. In order to implement the MSLRT, we need to compute the SLRT statistic $R(\psi)$ and the factor $Q(\psi)$ given in equations (12) and (16), respectively, in the paper. Once the MLE $(\widehat{\psi}, \widehat{\sigma}^2)$ and the constrained MLE $\widehat{\sigma}^2_\psi$ are obtained, the computation of $R(\psi)$ is straightforward, using the expression given in (12). Here we shall explain the computation of $Q(\psi)$.

Our derivations are based on the theory described in Wong and Wu (2000); we refer to this article for motivation and further details. In fact the expressions given below are obtained by simplifying the quantities in equations (7)−(17) in Wong and Wu (2000).

Let

$$\theta = (\mu, \sigma^2), \quad \text{and} \quad \psi = \psi(\theta) = \mu + \frac{\sigma^2}{2}.$$

For a function $f(\theta)$, we shall use the notation $f_\theta(\theta)$ to denote the 2×1 vector of first derivatives, and $f_{\theta\theta}(\theta)$ to denote the 2×2 matrix of second derivatives. The likelihood function, as a function of θ, can be expressed as

$$l(\theta) = k \ln \Phi\left(\frac{DL - \mu}{\sigma}\right) - (n-k)\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n-k}(X_i - \mu)^2.$$

The MLEs $\widehat{\mu}$ and $\widehat{\sigma}$ are the solutions to the equations

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \mu} &= -\frac{k}{\sigma}\frac{\phi(\xi)}{\Phi(\xi)} + \frac{(n-k)(\bar{X}_l - \mu)}{\sigma^2} = 0 \\ \frac{\partial l(\theta)}{\partial \sigma} &= -\frac{k\xi\phi(\xi)}{\sigma\Phi(\xi)} - \frac{n-k}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n-k}(X_i - \mu)^2 = 0, \end{aligned} \tag{A.1}$$

where $\xi = \frac{x_0 - \mu}{\sigma}$, as defined in Section 2.1, x_0 being the censoring threshold. Furthermore, the Fisher information matrix $J_\theta(\theta)$ is

$$J_\theta(\theta) = \begin{pmatrix} -\frac{\partial^2 l(\theta)}{\partial \mu^2} & -\frac{\partial^2 l(\theta)}{\partial \mu \partial \sigma} \\ -\frac{\partial^2 l(\theta)}{\partial \mu \partial \sigma} & -\frac{\partial^2 l(\theta)}{\partial \sigma^2} \end{pmatrix}, \tag{A.2}$$

where

$$\begin{aligned} -\frac{\partial^2 l(\mu,\sigma)}{\partial \mu^2} &= \frac{k}{\sigma^2} w(\xi)[w(\xi) + \xi] + \frac{n-k}{\sigma^2} \\ -\frac{\partial^2 l(\mu,\sigma)}{\partial \mu \partial \sigma} &= \frac{kw(\xi)}{\sigma^2}[\xi(\xi + w(\xi)) - 1] + \frac{2(n-k)(\bar{X}_l - \mu)}{\sigma^3} \\ -\frac{\partial^2 l(\mu,\sigma)}{\partial \sigma^2} &= \frac{k\xi w(\xi)}{\sigma^2}[\xi(\xi + w(\xi)) - 2] - \frac{n-k}{\sigma^2} + \frac{3(n-k)}{\sigma^4}(S_l^2 + (\bar{X}_l - \mu)^2), \end{aligned} \tag{A.3}$$

and $w(\xi) = \phi(\xi)/\Phi(\xi)$. Let $\hat{\theta} = (\hat{\mu}, \hat{\sigma})$ denote the MLE, and write

$$V = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \frac{X_1 - \widehat{\mu}}{\widehat{\sigma}} & \frac{X_2 - \widehat{\mu}}{\widehat{\sigma}} & \dots & \frac{X_{n-k} - \widehat{\mu}}{\widehat{\sigma}} \end{pmatrix} = \begin{pmatrix} V_{11} & V_{12} & \dots & V_{1,n-k} \\ V_{21} & V_{22} & \dots & V_{2,n-k} \end{pmatrix}.$$

Define

$$\begin{aligned} \phi_1(\theta) &= \sum_{j=1}^{n-k} \frac{\partial l(\theta)}{\partial X_j} V_{ij} = -\frac{n-k}{\sigma^2}(\bar{X}_l - \mu) \\ \phi_2(\theta) &= \sum_{j=1}^{n-k} \frac{\partial l(\theta)}{\partial X_j} V_{2j} = -\frac{(n-k)}{\widehat{\sigma}\sigma^2}[S_l^2 + (\bar{X}_l - \widehat{\mu})(\bar{X}_l - \mu)], \end{aligned} \tag{A.4}$$

and

$$\begin{aligned} \phi_\theta(\theta) &= \begin{pmatrix} \frac{\partial \phi_1(\theta)}{\partial \mu} & \frac{\partial \phi_1(\theta)}{\partial \sigma} \\ \frac{\partial \phi_2(\theta)}{\partial \mu} & \frac{\partial \phi_2(\theta)}{\partial \sigma} \end{pmatrix} \\ &= \begin{pmatrix} \frac{n-k}{\sigma^2} & \frac{2(n-k)}{\sigma^3}(\bar{X}_l - \mu) \\ \frac{n-k}{\sigma^2\widehat{\sigma}}(\bar{X}_l - \widehat{\mu}) & \frac{2(n-k)}{\sigma^3\widehat{\sigma}}[S_l^2 + (\bar{X}_l - \mu)(\bar{X}_l - \widehat{\mu})] \end{pmatrix}. \end{aligned} \tag{A.5}$$

For $\phi_\theta(\theta)$ defined above, we note that

$$\begin{aligned} |\phi_\theta(\theta)| &= \frac{2(n-k)^2 S_l^2}{\widehat{\sigma}\sigma^5} \\ \phi_\theta^{-1}(\theta) &= \frac{\sigma^2}{(n-k)S_l^2} \begin{pmatrix} [S_l^2 + (\bar{X}_l - \widehat{\mu})(\bar{X}_l - \mu)] & -\widehat{\sigma}(\bar{X}_l - \mu) \\ -\frac{\sigma(\bar{X}_l - \widehat{\mu})}{2} & \frac{\widehat{\sigma}\sigma}{2} \end{pmatrix}. \end{aligned} \tag{A.6}$$

Let $\psi(\theta) = \mu + \frac{\sigma^2}{2}$ and $\psi_\theta(\theta) = \left(\frac{\partial \psi(\theta)}{\partial \mu}, \frac{\partial \psi(\theta)}{\partial \sigma}\right) = (1, \sigma)$ so that

$$\psi_{\theta\theta}(\theta) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The constrained MLE of θ, when $\psi(\theta)$ is kept fixed at the value ψ_0, can be obtained by maximizing

$$K(\theta,\alpha) = l(\theta) + \alpha(\psi(\theta) - \psi_0)$$

with respect to α and θ. Let $\widehat{\alpha}$ and $\widehat{\theta}_{\psi_0} = (\widehat{\mu}_{\psi_0}, \widehat{\sigma}_{\psi_0})$ maximize $K(\theta,\alpha)$. We note that $\widehat{\sigma}^2_{\psi_0}$ satisfies equation (13) in the paper, and $\widehat{\mu}_{\psi_0} = \psi_0 - \widehat{\sigma}^2_{\psi_0}/2$. Equating the first derivative of $K(\theta,\alpha)$, with respect to μ, to zero, we also get

$$\widehat{\alpha} = \frac{kw(\widehat{\xi}_0)}{\widehat{\sigma}_{\psi_0}} - (n-k)\frac{(\bar{X}_l - \widehat{\mu}_{\psi_0})}{\widehat{\sigma}^2_{\psi_0}}, \quad \text{with} \quad \widehat{\xi}_0 = \frac{DL - \widehat{\mu}_{\psi_0}}{\widehat{\sigma}_{\psi_0}}.$$

The development of the MSLRT statistic also requires the information matrix, say $\tilde{J}_\theta(\theta)$, based on the "likelihood"

$$\tilde{l}(\theta) = l(\theta) + \widehat{\alpha}(\psi(\theta) - \psi_0).$$

It is clear that

$$\tilde{J}_\theta(\theta) = J_\theta(\theta) - \widehat{\alpha}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

where $J_\theta(\theta)$ is given in (A.2). Define

$$\begin{aligned}\chi(\theta) &= (1, \widehat{\sigma}_{\psi_0})\, \phi_\theta^{-1}(\widehat{\theta}_{\psi_0}) \begin{pmatrix} \phi_1(\theta) \\ \phi_2(\theta) \end{pmatrix} \\ &= \frac{\widehat{\sigma}^2_{\psi_0}}{\sigma^2}\left(\mu - \widehat{\mu}_{\psi_0} - \frac{\widehat{\sigma}^2_{\psi_0}}{2}\right), \end{aligned} \tag{A.7}$$

and

$$\chi(\widehat{\theta}) - \chi(\widehat{\theta}_{\psi_0}) = \frac{\widehat{\sigma}^2_{\psi_0}}{\widehat{\sigma}^2}\left(\widehat{\mu} - \widehat{\mu}_{\psi_0} - \frac{\widehat{\sigma}^2_{\psi_0}}{2}\right) + \frac{\widehat{\sigma}^2_{\psi_0}}{2}. \tag{A.8}$$

To get the above expression, we used the inverse matrix given in (A.6). Define

$$\widehat{\sigma}^2_\chi = (1, \widehat{\sigma}_{\psi_0})\, \tilde{J}_\theta^{-1}(\widehat{\theta}_{\psi_0}) \begin{pmatrix} 1 \\ \widehat{\sigma}_{\psi_0} \end{pmatrix}$$

$$= \frac{J_{\theta,22}(\widehat{\theta}_{\psi_0}) - \widehat{\alpha} - 2J_{\theta,12}(\widehat{\theta}_{\psi_0})\widehat{\sigma}_{\psi_0} + J_{\theta,11}(\widehat{\theta}_{\psi_0})\widehat{\sigma}^2_{\psi_0}}{|\tilde{J}_\theta(\widehat{\theta}_{\psi_0})|} \tag{A.9}$$

where $J_{\theta,ij}(\theta)$ is the (i,j) element of $J_\theta(\theta)$. Then

$$\widehat{\sigma}^2_\chi|\tilde{J}_\theta(\widehat{\theta}_{\psi_0})| = J_{\theta,22}(\widehat{\theta}_{\psi_0}) - \widehat{\alpha} - 2J_{\theta,12}(\widehat{\theta}_{\psi_0})\widehat{\sigma}_{\psi_0} + J_{\theta,11}(\widehat{\theta}_{\psi_0})\widehat{\sigma}^2_{\psi_0}. \tag{A.10}$$

Now let

$$\begin{aligned} Q(\psi_0) &= \operatorname{sign}\left(\widehat{\psi} - \psi_0\right) \left|\chi(\widehat{\theta}) - \chi(\widehat{\theta}_{\psi_0})\right| \left\{ \frac{\left|J_\theta(\widehat{\theta})\right| \left|\phi_\theta(\widehat{\theta}_{\psi_0})\right|^2}{\widehat{\sigma}^2_\chi \left|\tilde{J}_\theta(\widehat{\theta}_{\psi_0})\right| \left|\phi_\theta(\widehat{\theta})\right|^2} \right\}^{\frac{1}{2}} \\ &= \operatorname{sign}(\widehat{\psi} - \psi_0) \left| \frac{\widehat{\sigma}^2_{\psi_0}}{\widehat{\sigma}^2} \left(\widehat{\mu} - \widehat{\mu}_{\psi_0} - \frac{\widehat{\sigma}^2_{\psi_0}}{2}\right) + \frac{\widehat{\sigma}^2_{\psi_0}}{2} \right| \\ &\times \frac{\left|J_\theta(\widehat{\theta})\right|^{\frac{1}{2}} (\widehat{\sigma}/\widehat{\sigma}_{\psi_0})^5}{\left\{J_{\theta,22}(\widehat{\theta}_{\psi_0}) - \widehat{\alpha} - 2J_{\theta,12}(\widehat{\theta}_{\psi_0})\widehat{\sigma}_{\psi_0} + J_{\theta,11}(\widehat{\theta}_{\psi_0})\widehat{\sigma}^2_{\psi_0}\right\}^{1/2}}, \end{aligned} \tag{A.11}$$

where $J_{\theta,ij}(\hat{\theta}_{\psi_0})$ is the (i,j) element of $J_\theta(\hat{\theta}_{\psi_0})$. The quantity $Q(\psi)$ used in Section 3.2.3 of the paper is given by the above expression.

A2. The MSLRT statistic for a lognormal quantile

Here we shall give details of the derivation of the quantity $Q(\eta)$ given in Section 4.2.2. With $\theta = (\mu, \sigma)$, the parameter of interest is $\eta = \eta(\theta) = \mu + z_p\sigma$. Let $\hat{\theta} = (\widehat{\mu}, \widehat{\sigma})$ denote the MLE of θ. Then the MLE of η is $\widehat{\eta} = \widehat{\mu} + z_p\widehat{\sigma}$. Let $\widehat{\sigma}^2_\eta$ denote the constrained MLE of σ^2, obtained by solving equation (21) in the paper. Then the constrained MLE of θ, say $\widehat{\theta}_\eta$, is given by

$$\widehat{\theta}_\eta = (\eta - z_p\widehat{\sigma}_\eta, \widehat{\sigma}_\eta) = (\widehat{\mu}_\eta, \widehat{\sigma}_\eta).$$

The constrained MLE can also be obtained by maximizing $l(\eta, \sigma^2) + \alpha(\eta(\theta) - \eta_0)$, where α is a lagrange multiplier and η_0 is a fixed value of $\eta = \eta(\theta)$. As before, the development of the MSLRT statistic also requires the information matrix, say $\tilde{J}_\theta(\theta)$, based on the "likelihood" $\tilde{l}(\eta, \sigma^2) = l(\eta, \sigma^2) + \widehat{\alpha}(\eta(\theta) - \eta_0)$, $\widehat{\alpha}$ being the value of α that maximizes $l(\eta, \sigma^2) + \alpha(\eta(\theta) - \eta_0)$. Since $\eta(\theta) = \mu + z_p\sigma$, it is easy to verify that $\tilde{J}_\theta(\theta) = J_\theta(\theta)$, the information matrix based on $l(\eta, \sigma^2)$. The latter information matrix is given in equations (A.2) and (A.3).

Define $(\phi_1(\theta), \phi_2(\theta))$ and $\phi_\theta(\theta)$ as in (A.4) and (A.5), respectively, and let

$$\chi(\theta) = (1, z_p)\, \phi_\theta^{-1}(\widehat{\theta}_\eta) \begin{pmatrix} \phi_1(\theta) \\ \phi_2(\theta) \end{pmatrix} = \frac{\widehat{\sigma}_\eta^2}{\sigma^2}\left(\mu - \widehat{\mu}_\eta - \frac{\widehat{\sigma}_\eta z_p}{2}\right). \tag{A.12}$$

It is easy to see that

$$\chi(\widehat{\theta}) - \chi(\widehat{\theta}_\eta) = \frac{\widehat{\sigma}_\eta^2}{\widehat{\sigma}^2}\left(\widehat{\mu} - \widehat{\mu}_\eta - \frac{\widehat{\sigma}_\eta z_p}{2}\right) + \frac{\widehat{\sigma}_\eta z_p}{2}. \tag{A.13}$$

Also define

$$\widehat{\sigma}_\chi^2 = (1, z_p)\, \tilde{J}_\theta^{-1}(\widehat{\theta}_\eta) \begin{pmatrix} 1 \\ z_p \end{pmatrix} = (1, z_p)\, J_\theta^{-1}(\widehat{\theta}_\eta) \begin{pmatrix} 1 \\ z_p \end{pmatrix}, \tag{A.14}$$

where we have used the property $\tilde{J}_\theta(\theta) = J_\theta(\theta)$. With $|\phi_\theta(\theta)|$ as given in equation (A.6), it is easy to see that $|\phi_\theta(\widehat{\theta}_\eta)|/|\phi_\theta(\widehat{\theta}_\eta)| = \widehat{\sigma}^5/\widehat{\sigma}_\eta^5$. Now define

$$\begin{aligned} Q(\eta) &= \operatorname{sign}(\widehat{\eta} - \eta)\left|\chi(\widehat{\theta}) - \chi(\widehat{\theta}_\eta)\right| \times \left\{ \frac{\left|J_\theta(\widehat{\theta})\right|\left|\phi_\theta(\widehat{\theta}_\eta)\right|^2}{\widehat{\sigma}_\chi^2\left|\tilde{J}_\theta(\widehat{\theta}_\eta)\right|\left|\phi_\theta(\widehat{\theta})\right|^2} \right\}^{\frac{1}{2}} \\ &= \operatorname{sign}(\widehat{\eta} - \eta)\left|\frac{\widehat{\sigma}_\eta^2}{\widehat{\sigma}^2}\left(\widehat{\mu} - \widehat{\mu}_\eta - \frac{\widehat{\sigma}_\eta z_p}{2}\right) + \frac{\widehat{\sigma}_\eta z_p}{2}\right| \times \left\{ \frac{1}{\widehat{\sigma}_\chi^2} \times \left(\frac{\widehat{\sigma}^2}{\widehat{\sigma}_\eta^2}\right)^5 \right\}^{\frac{1}{2}}. \end{aligned} \tag{A.15}$$

The quantity $Q(\eta)$ used in Section 4.2.2 of the paper is given by the above expression.

REFERENCE

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