2018

Two-sided tolerance intervals for members of the (log)-location-scale family of distributions

Miao Yuan
Virginia Tech

Yili Hong
Virginia Tech

Luis A. Escobar
Louisiana State University

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

Two-sided tolerance intervals for members of the (log)-location-scale family of distributions

Abstract

In this paper, we propose methods to calculate exact factors for two-sided control-the-centre and control-both-tails tolerance intervals for the (log)-location-scale family of distributions, based on complete or Type II censored data. With Type I censored data, exact factors do not exist. For this case, we developed an algorithm to compute approximate factors. Our approaches are based on Monte Carlo simulations. We also provide algorithms for computing TIs that control the probability in both tails of a distribution. A simulation study for Type I censored data shows that the estimated coverage probability is close to the nominal confidence level when the expected number of uncensored observations is moderate to large. We illustrate the methods with applications using different combinations of distributions and types of censoring.

Keywords

Censored data, lognormal, maximum likelihood, Monte-Carlo simulation, pivotal quantity, Weibull

Disciplines

Statistics and Probability

Comments

This is an Accepted Manuscript of an article published by Taylor & Francis as Yuan, Miao, Yili Hong, Luis A. Escobar, and William Q. Meeker. "Two-sided tolerance intervals for members of the (log)-location-scale family of distributions." *Quality Technology & Quantitative Management* 15, no. 3 (2018): 374-392. DOI: 10.1080/16843703.2016.1226594. Posted with permission.

Two-sided Tolerance Intervals for Members of the (Log)-Location-Scale Family of Distributions

Miao Yuan and Yili Hong
Department of Statistics
Virginia Tech
Blacksburg, VA 24061

Luis A. Escobar
Department of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803

William Q. Meeker
Department of Statistics
Iowa State University
Ames, IA 50011

Abstract

In this paper, we propose methods to calculate exact factors for two-sided control-the-center and control-both-tails tolerance intervals (TI) for the (log)-location-scale family of distributions, based on complete or Type II censored data. With Type I censored data, exact factors do not exist. For this case we developed an algorithm to compute approximate factors. Our approaches are based on Monte-Carlo simulations. We also provide algorithms for computing TIs that control the probability in both tails of a distribution. A simulation study for Type I censored data shows that the estimated coverage probability (CP) is close to the nominal confidence level when the expected number of uncensored observations is moderate to large. We illustrate the methods with applications using different combinations of distributions and types of censoring.

Key Words: Censored data, Lognormal, Maximum likelihood, Monte-Carlo simulation, Pivotal quantity, Weibull.

1 Introduction

1.1 Motivation

In many applications such as quality control and reliability engineering, one is often interested in estimating the content of a distribution with a high degree of confidence, based on an observed random sample (perhaps censored) from that distribution. Such an estimation procedure is different from obtaining a confidence interval to contain a scalar characteristic, such as a distribution mean or quantile. For example, one often needs an interval that can cover at least a proportion β of a distribution. Tolerance intervals (TIs) are frequently used to describe the distribution of a process output, a particular random quantity characteristic, or the lifetime distribution of a product (e.g., Krishnamoorthy and Mathew 2009). More applications of TIs are also described in Hahn and Meeker (1991). A TI to contain a proportion of the distribution is used for this purpose. In this paper, we use $(\beta, 1-\alpha)$ to denote a TI that has $100(1-\alpha)\%$ confidence to cover at least a content β of a population.

Members of the (log)-location-scale family of distributions (e.g., the normal, the Weibull and the lognormal) are commonly used to model data in the physical and engineering sciences. Applications involving life testing often result in right-censored data. Thus, the objective of this paper is to develop a general procedure to compute two-sided TI for the location-scale and log-location-scale families of distributions based on complete and right censored data.

1.2 Literature Review

In this paper, we consider TIs for continuous distributions. For the normal distribution, TIs based on both complete and censored data have been well developed. Specifically, Odeh and Owen (1980) provided exact factors of one-sided tolerance bounds (TBs) and two-sided TIs with various combinations of confidence levels, contents, and sample sizes for a normal distribution. The tables of Odeh and Owen (1980), however, only apply to the case where the degrees of freedom is equal to the sample size minus one. Weissberg and Beatty (1960) and Howe (1969) proposed approximate factors for normal TIs with sample variance having arbitrary degrees of freedom. Eberhardt, Mee, and Reeve (1989) developed a FORTRAN program to compute the exact factors of normal TIs for a wide ranges of degrees of freedom, effective sample sizes, and confidence levels. Their methods show advantages especially when the effective sample size is much smaller than the degree of freedom, for example, in a regression setting.

In the situation where censored observations are present, Krishnamoorthy and Mathew

(2009) discussed the calculation of one-sided TBs for normal and the Weibull distributions. One-sided TBs for lognormal and the smallest extreme value (SEV) distribution can be obtained based on their one-to-one relationship with normal and the Weibull distributions, respectively. For the Laplace distribution, computation of one-sided TBs and two-sided TIs were studied, for example, in Shyu and Owen (1986a) and Shyu and Owen (1986b), using Monte-Carlo simulations. For symmetric distributions in the location-scale family, Krishnamoorthy and Xie (2011) derived algorithms for computing control-the-center TIs and equal-tail TIs using pivotal quantities based on ML estimators under complete and Type II censored data. They also considered an adjusted method for Type I censored data. Our objective is to provide a generic algorithm that can compute TIs for both symmetric and non-symmetric distributions in the (log)-location-scale family. For the (log)-location-scale family of distributions, Xie et al. (2014) developed a general method to calculate the exact one-sided prediction bounds and two-sided prediction intervals under complete and Type II censored data, and an adjusted procedure for Type I censored data.

Bergquist (2006) developed a parametric bootstrap procedure to calculate the one-sided TBs and two-sided TIs for response of a (non)-linear-mixed-effect model in the pharmaceutical application setting. They concluded that the bootstrap TIs usually have considerably lower actual CP than the nominal confidence level when the sample size is small to moderate, in the situations involving a simple random sample and a (non)-linear-mixed-effect model. The performance of their parametric bootstrap procedures largely depends on the accuracy of the model parameter estimates. Our methods, however, are based on the pivotal properties of the location-scale family of distributions. Thus, the computed TI factors do not depend on the parameter estimates, when data are complete or Type II censored. Besides, we evaluate and control the CP directly in the Monte-Carlo simulations to calculate the factors. Hence our methods do not have the problems pointed out in Bergquist (2006).

One-sided TBs have been well studied in literature. The lower (upper) TB is one-sided lower (upper) confidence bound with confidence level $100(1-\alpha)\%$ on the β quantile [$(1-\beta)$ quantile]. Our focus will therefore be on two-sided TIs. There are two types of commonly-used TIs: control-the-center TIs and control-both-tails TIs. One can also see description of these two types of TI in Odeh and Owen (1980), Hahn and Meeker (1991, Chapter 4) and Krishnamoorthy and Mathew (2009, Chapter 1).

1.3 Overview

The rest of this paper is organized as follows. Section 2 introduces the data and distributions used in this paper. Section 3 introduces definitions of two-sided TI procedures for the (log)-

location-scale family of distributions. Section 4 describes algorithms for computing TIs for the (log)-location-scale family of distributions. Section 5 focuses on applications of the developed procedure for three examples. Section 6 discusses the results of a simulation study for Type I censored data. Section 7 contains conclusions and some areas for future research.

2 Data and Model

2.1 Data

We consider TIs for complete, Type I (time), and Type II (failure) censored data. For complete and Type II censored data, our TI procedure is exact. For Type I censored data, we derive an approximate procedure.

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ denote n independent random variables corresponding to the response of interest. The censoring indicator δ_i, $i = 1, 2, \ldots, n$ is defined as $\delta_i = 1$ when X_i is uncensored and $\delta_i = 0$ when X_i is right censored. The observed data are denoted by $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)$. For simplicity, and without loss of generality, we assume that $x_1 < \cdots < x_n$.

We consider three types of data:

1. Complete data. In this case the $x_1, x_2, \ldots, x_n$ are the realizations of the X_i. Because there is no censoring, $\delta_i = 1$ for all i.

2. Type II censored data. In this case the data consist of the r smallest observations $x_1 < \cdots < x_r$ in the sample of size n and the additional information that the remaining $(n - r)$ sample values are censored at x_r. Then $\delta_i = 1$ for $i = 1, \ldots, r$ and $\delta_i = 0$ for $i = r+1, \ldots, n$. Note that r is a pre-specified integer between 1 and n but x_r is random. To ensure estimability of the parameters we consider situations in which $r \geq 2$. Type II censoring sometimes arises in life testing.

3. Type I censored data. The data consist of the r smallest observations $(x_1, \ldots, x_r)$ that satisfy $x_i \leq x_c$, where the fixed bound x_c is pre-specified, and the additional information that the remaining $(n - r)$ observations are censored at x_c. Then $\delta_i = 1$ for $i = 1, \ldots, r$ and $\delta_i = 0$ for $i = r + 1, \ldots, n$. Note that, in this case, the number of uncensored observations r is random and x_c is fixed. We exclude the case $r = 0$ because the maximum likelihood (ML) estimator of the parameters does not exist. Type I censored data arise in life testings and also in other applications such as when a measurement system has an upper bound for its readings.

2.2 Model

Because the (log)-location-scale family of distributions has many applications for modeling lifetime data or time to event data in reliability and survival applications, and many other areas of science and engineering, our study focuses on computing TIs for members in this family of distributions. The location-scale family of distributions is characterized by a location parameter $-\infty < \mu < \infty$ and a scale parameter $\sigma > 0$, which are typically unknown and need to be estimated from data. The cumulative distribution function (cdf) and the probability density function (pdf) of the location-scale distributions are

$$F(x;\boldsymbol{\theta}) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \text{and} \quad f(x;\boldsymbol{\theta}) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right), \tag{1}$$

respectively, where $\boldsymbol{\theta} = (\mu, \sigma)$. A random variable Y belongs to the log-location-scale family if $X = \log(Y)$ follows a location-scale distribution. For the log-location-scale family, $\exp(\mu)$ is a scale parameter and $\sigma > 0$ is a shape parameter. The cdf and pdf of the log-location-scale family of distributions are

$$F(y;\boldsymbol{\theta}) = \Phi\left[\frac{\log(y)-\mu}{\sigma}\right] \quad \text{and} \quad f(y;\boldsymbol{\theta}) = \frac{1}{\sigma y}\phi\left[\frac{\log(y)-\mu}{\sigma}\right],$$

respectively. Here $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and pdf for a standard distribution in the location-scale family ($\mu = 0$ and $\sigma = 1$), respectively.

Some commonly-used distributions in the location-scale family, their standard pdf $\phi(x)$ and cdf $\Phi(x)$, along with the corresponding log-location-scale distributions are summarized as follows:

$$\begin{aligned}
\text{Normal (Lognormal)}: &\quad \phi_{\text{norm}}(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right), &\quad \Phi_{\text{norm}}(z) &= \int_{-\infty}^{z}\phi_{\text{norm}}(v)dv \\
\text{Logistic (Loglogistic)}: &\quad \phi_{\text{logis}}(z) = \frac{\exp(z)}{[1+\exp(z)]^2}, &\quad \Phi_{\text{logis}}(z) &= \frac{\exp(z)}{1+\exp(z)} \\
\text{LEV (Fréchet)}: &\quad \phi_{\text{sev}}(z) = \exp[-z-\exp(-z)], &\quad \Phi_{\text{sev}}(z) &= \exp[-\exp(-z)] \\
\text{SEV (Weibull)}: &\quad \phi_{\text{lev}}(z) = \exp[z-\exp(z)], &\quad \Phi_{\text{lev}}(z) &= 1-\exp[-\exp(z)].
\end{aligned}$$

where SEV and LEV are the smallest extreme value and largest extreme value distributions, respectively.

2.3 Parameter Estimation

Because parametric distributions in the (log)-location-scale family are used to model the (possibly censored) data, it is desirable to obtain parameter estimates by the maximum likelihood (ML) method. The ML method is easy to implement given a parametric model and the data censoring type.

Assuming the right censored samples are independent and identically distributed (iid), the likelihood function of the parameters $\boldsymbol{\theta} = (\mu, \sigma)$ given the data is,

$$L(\boldsymbol{\theta}|\text{DATA}) = \prod_{i=1}^{n} [f(x_i)]^{\delta_i} [1 - F(x_c)]^{1-\delta_i},$$

where $f(\cdot)$ and $F(\cdot)$ are the pdf and cdf given in (1), respectively. Because ML estimates of the (log)-location-scale family generally do not have explicit forms for censored data, numerical methods are applied to find the values of $\boldsymbol{\theta}$ that maximize the log likelihood function. The ML estimates are denoted by $\widehat{\boldsymbol{\theta}} = (\widehat{\mu}, \widehat{\sigma})$.

3 Two-sided Tolerance Interval Procedures

In this paper, we develop methods to construct two-sided TIs for the members in the (log)-location-scale family of distributions. In particular, we construct a $(\beta, 1-\alpha)$ TI,

$$\left[\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha), \quad \widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha)\right] \tag{2}$$

for given data $\boldsymbol{x}$ from a particular member of this family of distributions. Here, β is the content of the distribution that the TI should cover with $100(1-\alpha)\%$ confidence. We focus on the TIs for the location-scale family of distributions. For a distribution in the log-location-scale family, we first compute, using the log transformed data, the TI for the corresponding location-scale distribution as $\left[\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha), \quad \widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha)\right]$. Then, the desired TI is obtained as

$$\left\{\exp\left[\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha)\right], \quad \exp\left[\widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha)\right]\right\}.$$

3.1 Control-the-Center TIs

Let $F(x; \boldsymbol{\theta})$ be the cdf of the random variable X from which the not-yet-observed data $\boldsymbol{X}$ will be taken. For notational simplicity, let

$$\Delta F_L(\boldsymbol{x}) = F\left[\widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha); \boldsymbol{\theta}\right] \quad \text{and} \quad \Delta F_U(\boldsymbol{x}) = 1 - F\left[\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha); \boldsymbol{\theta}\right] \tag{3}$$

be the population contents below $\widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha)$ and above $\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha)$ for any realization of $\boldsymbol{X} = \boldsymbol{x}$, respectively. Then, the population content of the TI in (2) is

$$\Delta F(\boldsymbol{x}) = 1 - \Delta F_U(\boldsymbol{x}) - \Delta F_L(\boldsymbol{x}). \tag{4}$$

In applications, one is interested in the probability with which the random quantity $\Delta F(\boldsymbol{X})$ exceeds β. The TI in (2) is said to be an exact $(\beta, 1-\alpha)$ control-the-center TI if

$$\Pr\{\Delta F(\boldsymbol{X}) > \beta\} = 1 - \alpha. \tag{5}$$

The coverage probability (CP) of a control-the-center TI is

$$\mathrm{CP}(\boldsymbol{\theta}) = \Pr\{\Delta F(\boldsymbol{X}) > \beta\} = \mathrm{E}_{\boldsymbol{X}}\left(\mathrm{I}\{\Delta F(\boldsymbol{X}) > \beta\}\right), \tag{6}$$

where $\mathrm{E}_{\boldsymbol{X}}$ is the expectation with respect to the distribution of $\boldsymbol{X}$, and the indicator function $\mathrm{I}[A]$ is equal to 1 when the statement A is true and is equal to 0 otherwise. Note that in some cases, the CP will depend on the parameter $\boldsymbol{\theta}$ through the probability distribution of the data $\boldsymbol{X}$.

For members of the location-scale family of distributions, the TI is constructed using the following forms,

$$\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha) = \widehat{\mu} + g_{L(1-\alpha,\beta)}\widehat{\sigma} \quad \text{and} \quad \widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha) = \widehat{\mu} + g_{U(1-\alpha,\beta)}\widehat{\sigma}, \tag{7}$$

where $g_L = g_{L(1-\alpha,\beta)}$ and $g_U = g_{U(1-\alpha,\beta)}$ are the factors that provide the desired CP.

Result 1 *Let $\boldsymbol{Z} = (Z_1, Z_2)$ where $Z_1 = (\widehat{\mu} - \mu)/\sigma$ and $Z_2 = \widehat{\sigma}/\sigma$. The CP in* (6) *only depends on the distribution of $\boldsymbol{Z}$. That is,*

$$\mathrm{CP}(\boldsymbol{\theta}) = E_{\boldsymbol{Z}}\left(\mathrm{I}\{\Phi(Z_1 + g_U Z_2) - \Phi(Z_1 + g_L Z_2) > \beta\}\right), \tag{8}$$

where the factors g_L and g_U are chosen to satisfy

$$1 - \alpha = E_{\boldsymbol{Z}}\left(\mathrm{I}\{\Phi(Z_1 + g_U Z_2) - \Phi(Z_1 + g_L Z_2) > \beta\}\right), \tag{9}$$

when data are complete or Type II censored.

The proof of **Result 1** is given in Appendix A. When data are complete or Type II censored, the distributions of pivotal quantities $\boldsymbol{Z}$ for a location-scale distribution do not depend on any unknown parameters (e.g., Krishnamoorthy and Mathew 2009, pages 16–17). When data are Type I censored, the quantities $\boldsymbol{Z}$ are approximately pivotal and thus their distributions depend on the unknown parameters. In particular, we have the following result for Type I censored data.

Result 2 *For any given censoring time x_c, the quantities $\boldsymbol{Z}$ are approximately pivotal under Type I censoring. Their distributions only depend on the expected fraction of uncensored observations $p_f = \Phi[(x_c - \mu)/\sigma]$.*

A proof of this result is given in Appendix B.

3.2 Control-Both-Tails TIs

The control-the-center TI only guaranties that the population content is at least β but the procedure is not specific on the amount of content in each tail. In some applications, in contrast to the control-the-center TI, one needs to construct a more stringent TI that leaves no more than a specified amount of content in each tail of the distribution. The control-both-tails TI is used for this purpose. See Hahn and Meeker (1991, Chapter 4) for a description of the two alternative types of TIs. Although it is possible to specify different probabilities in each tail, it is more common that the specified probabilities are equal and we will follow that convention. The TI in (2) is said to be an exact $(\beta, 1-\alpha)$ control-both-tails TI if

$$\Pr\left\{\Delta F_L(\boldsymbol{X}) \leq (1-\beta)/2, \Delta F_U(\boldsymbol{X}) \leq (1-\beta)/2\right\} = 1-\alpha. \tag{10}$$

Note that the population content in-between the interval endpoints is at least β for the TI in (10). The TI in (10), however, is more stringent than the TI in (5) because it requires the same content in each tail.

The CP of a control-both-tails TI is

$$\begin{aligned} \mathrm{CP}(\boldsymbol{\theta}) &= \Pr\left\{\Delta F_L(\boldsymbol{X}) \leq (1-\beta)/2, \Delta F_U(\boldsymbol{X}) \leq (1-\beta)/2\right\} \\ &= \mathrm{E}_{\boldsymbol{X}}\left(\mathrm{I}\left\{\Delta F_L(\boldsymbol{X}) \leq (1-\beta)/2, \Delta F_U(\boldsymbol{X}) \leq (1-\beta)/2\right\}\right). \end{aligned} \tag{11}$$

For members of the location-scale family of distributions, the TI is constructed using the following forms

$$\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha) = \widehat{\mu} + g'_{L(1-\alpha,\beta)}\widehat{\sigma} \quad \text{and} \quad \widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha) = \widehat{\mu} + g'_{U(1-\alpha,\beta)}\widehat{\sigma},$$

where $g'_L = g'_{L(1-\alpha,\beta)}$ and $g'_U = g'_{U(1-\alpha,\beta)}$ are the factors that provide the desired CP.

Result 3 *The CP in* (11) *only depends on the distribution of* $\boldsymbol{Z}$. *That is,*

$$\mathrm{CP}(\boldsymbol{\theta}) = E_{\boldsymbol{Z}}\left(\mathrm{I}\left\{\Phi(Z_1 + g'_L Z_2) \leq (1-\beta)/2, \Phi(Z_1 + g'_U Z_2) \geq (1+\beta)/2\right\}\right) \tag{12}$$

and the factors g'_L *and* g'_U *are chosen to satisfy*

$$1-\alpha = E_{\boldsymbol{Z}}\left(\mathrm{I}\left\{\Phi(Z_1 + g'_L Z_2) \leq (1-\beta)/2, \Phi(Z_1 + g'_U Z_2) \geq (1+\beta)/2\right\}\right), \tag{13}$$

when data are complete or Type II censored.

The proof of **Result 3** is similar to that of **Result 1**. Thus it is omitted.

3.3 TIs with Equal Error Probabilities

The pairs of (g_L, g_U) for a $(\beta, 1-\alpha)$ TI are not uniquely determined. It is appropriate to calculate a "balanced" TI, such that the two one-sided TBs constructed by its lower and upper bounds have equal error probabilities. That is, if $[\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha),\ \widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha)]$ is a balanced TI, then using the lower end point (the upper end point) of the TI as a lower (upper) tolerance bound to contain a fraction $1-\beta/2$ of the population has the desirable property that the coverage probabilities of these two bounds are the same. Specifically, we define the CP for each of these bounds as

$$\mathrm{CP}_U(\boldsymbol{\theta}) = \Pr\left\{F\left[\widetilde{T}(\boldsymbol{X}, \beta, 1-\alpha); \boldsymbol{\theta}\right] > \beta/2\right\} \tag{14}$$

$$\mathrm{CP}_L(\boldsymbol{\theta}) = \Pr\left\{1 - F\left[\underset{\sim}{T}(\boldsymbol{X}, \beta, 1-\alpha); \boldsymbol{\theta}\right] > \beta/2\right\}. \tag{15}$$

A $(\beta, 1-\alpha)$ TI with equal error probabilities is the TI in (5) or (10) with an additional constraint that

$$\mathrm{CP}_U(\boldsymbol{\theta}) = \mathrm{CP}_L(\boldsymbol{\theta}). \tag{16}$$

4 Computation of Tolerance Intervals

In this section, we develop algorithms to calculate factors of control-the-center and control-both-tails TI with CP equal to the nominal confidence level. To obtain the unique pair of factors, equation (9) and (16) need to be solved for control-the-center and (13) and (16) for control-both-tails TI.

Because the joint distribution of $\boldsymbol{Z}$ is complicated and does not have an explicit form, especially with censored observations, and there are usually no simplifications for (8) and (12) when the distribution in the location-scale family is non-symmetric, the equations need to be solved numerically. In our algorithm, the CP in (8) or (12) of a given procedure is evaluated directly through Monte-Carlo simulations from the joint distribution of $\boldsymbol{Z}$.

4.1 Control-the-Center TIs

For complete, Type II, or Type I censored data from the location-scale family of distribution, we develop **Algorithm 1** to calculate a control-the-center TI.

Algorithm 1

Denote by $\widehat{\mu}$ and $\widehat{\sigma}$ the ML estimators for μ and σ obtained from the available data.

1. Use the corresponding location-scale distribution with parameters $(\widehat{\mu}, \widehat{\sigma})$ to simulate $\boldsymbol{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ with the same sample size n and censoring pattern as the available data. For notation simplicity, the entries in $\boldsymbol{x}^*$ are sorted such that $x_1^* < x_2^* < \cdots < x_n^*$.
   - If the original data are complete, the data for the this iteration of the simulation are $x_1^*, \ldots, x_n^*$.
   - If the original data are Type II censored, then the data for the simulation are $x_1^*, \ldots, x_r^*$ and the additional information that the remaining $(n-r)$ data points are censored at x_r^*.
   - If the original data are Type I censored, retain as exact observations the r realizations $x_1^* < \cdots < x_r^*$ values that satisfy $x_i^* \le x_c$ where x_c is the same as the censoring bound in the data. The remaining $(n-r)$ simulated values are censored at x_c. In this censoring scheme r is random. Note that the expected proportion of non-censored observations, in the simulation, equal to the ML $\widehat{p}_f = \Phi[(x_c - \widehat{\mu})/\widehat{\sigma}]$ for the fraction of non-censored observations for the original data. Samples with $r = 0$ are discarded.
2. Compute the ML estimates $(\widehat{\mu}^*, \widehat{\sigma}^*)$ using the simulated data $\boldsymbol{x}^*$.
3. Repeat steps 1-2 B times. The simulated data in each iteration and the corresponding ML estimates are denoted as $\boldsymbol{x}_j^*$ and $(\widehat{\mu}_j^*, \widehat{\sigma}_j^*)$, respectively. Then for any pair of (g_L, g_U) in the two-dimensional domain, we calculate the CP using

$$\mathrm{CP}(g_L, g_U) \approx \frac{1}{B}\sum_{j=1}^{B} \mathrm{I}\Big[\Phi\left(\frac{\widehat{\mu}_j^* + g_U\widehat{\sigma}_j^* - \widehat{\mu}}{\widehat{\sigma}}\right) - \Phi\left(\frac{\widehat{\mu}_j^* + g_L\widehat{\sigma}_j^* - \widehat{\mu}}{\widehat{\sigma}}\right) > \beta\Big].$$

In step 3, one specifies ranges for g_L and g_U that are typically wide enough. Then one computes the CP for a grid of g_L and g_U. The factors (g_L, g_U) that provide $\mathrm{CP}(g_L, g_U) = 1-\alpha$ are chosen subject to the equal error probability in each tail constraint introduced in Section 3.3. The intersection point of the two curves (17) and (18) are calculated using linear interpolation.

$$\mathrm{CP}(g_L, g_U) = 1 - \alpha, \tag{17}$$

$$\mathrm{CP}_L(g_L) = \mathrm{CP}_U(g_U). \tag{18}$$

Computing the line corresponding to the constraint (18) requires calculation of the CP for one-sided TBs as follows

$$\mathrm{CP}_U(g_U) = \mathrm{CP}_U\left[\widetilde{T}(\boldsymbol{x}, \beta, 1-\alpha)\right] \approx \frac{1}{B}\sum_{j=1}^{B} \mathrm{I}\left[\Phi\left(\frac{\widehat{\mu}_j^* + g_U\widehat{\sigma}_j^* - \widehat{\mu}}{\widehat{\sigma}}\right) \geq \beta/2\right],$$

$$\mathrm{CP}_L(g_L) = \mathrm{CP}_L\left[\underset{\sim}{T}(\boldsymbol{x}, \beta, 1-\alpha)\right] \approx \frac{1}{B}\sum_{j=1}^{B} \mathrm{I}\left[\Phi\left(\frac{\widehat{\mu}_j^* + g_L\widehat{\sigma}_j^* - \widehat{\mu}}{\widehat{\sigma}}\right) \leq 1-\beta/2\right],$$

where $\widehat{\mu}_j^*, \widehat{\sigma}_j^*$ are the ML estimates used to compute $\mathrm{CP}(g_L, g_U)$.

When data are complete or Type II censored, the quantities Z_1 and Z_2 have the exact pivotal properties. Thus, $\widehat{\mu} = 0$ and $\widehat{\sigma} = 1$ can also be used in **Algorithm 1**, providing the advantage that for a specified distribution in the location-scale family, the factors (g_L, g_U) for given n and $r \leq n$ could be computed once and for all. Similarly, $\widehat{\mu} = 0$ and $\widehat{\sigma} = 1$ can be specified for the Monte-Carlo simulations when calculating CP of the one-sided TBs for the equal-tail constraint (18).

When data are Type I censored, however, Z_1 and Z_2 are only approximately pivotal. Asymptotically, the joint distribution of $\boldsymbol{Z}$ only depends on the expected fraction of uncensored observations p_f, which can be estimated using the ML method as $\widehat{p}_f = (x_c - \widehat{\mu})/\widehat{\sigma}$. Hence, $\widehat{\mu}$ and $\widehat{\sigma}$ are specified in **Algorithm 1**, providing the advantage that the censoring point x_c in the Monte-Carlo simulations is the same as that of the real data.

4.2 Control-Both-Tails TIs

For distributions in the location-scale family, we can use a procedure that is similar to the control-the-center TI procedure to construct control-both-tails TI with equal tail-probabilities. In this case, CP in **Algorithm 1** should be evaluated as:

$$\mathrm{CP}(g_L', g_U') \approx \frac{1}{B}\sum_{j=1}^{B} \mathrm{I}\left[\Phi\left(\frac{\widehat{\mu}_j^{**} + g_U'\widehat{\sigma}_j^{**} - \widehat{\mu}}{\widehat{\sigma}}\right) \geq \frac{1+\beta}{2} \text{ and } \Phi\left(\frac{\widehat{\mu}_j^{**} + g_L'\widehat{\sigma}_j^{**} - \widehat{\mu}}{\widehat{\sigma}}\right) \leq \frac{1-\beta}{2}\right].$$

The contour line $\mathrm{CP}_L(g_L') = \mathrm{CP}_U(g_U')$ of factors for equal-tail TIs is the same, regardless of the types of TI (i.e., control-the-center or control-both-tails).

4.3 Computation of TIs for the Log-Location-Scale Family of Distributions

If data follow a distribution in the log-location-scale family, we first take the log transformation of the data. Then for the transformed data from the corresponding location-scale

Table 1: Levels of Lead in Air ($\mu g/m^3$)

200	120	15	7	8	6	48	61
380	80	29	1000	350	1400	110	

distribution, **Algorithm 1** can be implemented to compute the factors of the desired TI. Last, take anti-logs (exponential) of the TI endpoints for the log transformed data and we get the TI for the original data as $\left[\exp(\widehat{\mu} + g_L\widehat{\sigma}), \quad \exp(\widehat{\mu} + g_U\widehat{\sigma})\right]$.

5 Applications

In this section, our method is illustrated in three different applications to compute TIs for the Weibull, the lognormal, and the loglogistic distributions.

5.1 Air Lead Level Data

Levels of lead-in-air data shown in Table 1 were studied in Krishnamoorthy, Mathew, and Ramachandran (2006). This dataset, collected by the National Institute of Occupational Safety and Health (NIOSH) at the Alma American Labs on February 23, 1989, contains 15 air lead levels from different spots within the facility. To evaluate the health risk for the staff, researchers needed to compute the ($\beta = 0.9, 1 - \alpha = 0.9$) two-sided TIs to describe the distribution of the air lead levels, based on for the complete data. Krishnamoorthy, Mathew, and Ramachandran (2006) suggested that the data can be described well by a lognormal distribution.

The ML estimates of lognormal distribution parameters are $\widehat{\mu} = 4.3329$ and $\widehat{\sigma} = 1.6805$. Following **Algorithm 1**, we simulate samples of size 15 from the standard normal distribution. The number of Monte-Carlo samples was chosen to be 100,000 to make simulation error negligible. Figure 2a is a contour plots of pairs of (g_L, g_U) that yield control-the-center TIs with content $\beta = 0.9$ and eight different levels of CP between 0.90 and 0.97. The equal-tail TI is illustrated at the confidence level 90%. The factors of an equal-tail TI to control the center are presented as the coordinates of the intersection point in Figure 2b. We calculate the coordinates of the intersection point of $\text{CP}(g_L, g_U) = 0.9$ with the contour line $\text{CP}_L(g_L) = \text{CP}_U(g_U)$ as described in Section 4.1. Figure 3a and 3b are similar contour plots for control-both-tails TIs. The results of the computed factors and TIs are summarized in Table 2.

That is, by the control-the-center TI, we have 90% confidence that a proportion 0.9 of

Table 2: Tolerance intervals for the levels of lead in air data

Model	ML Estimates $\widehat{\mu}$	$\widehat{\sigma}$	TI Control-Center	TI Control-Tails
Lognormal	4.3329	1.6805	$g_L = -2.37,\ g_U = 2.37$ TI = (1.42, 4087.48)	$g'_L = -2.61,\ g'_U = 2.61$ TI = (0.95 6118.09)

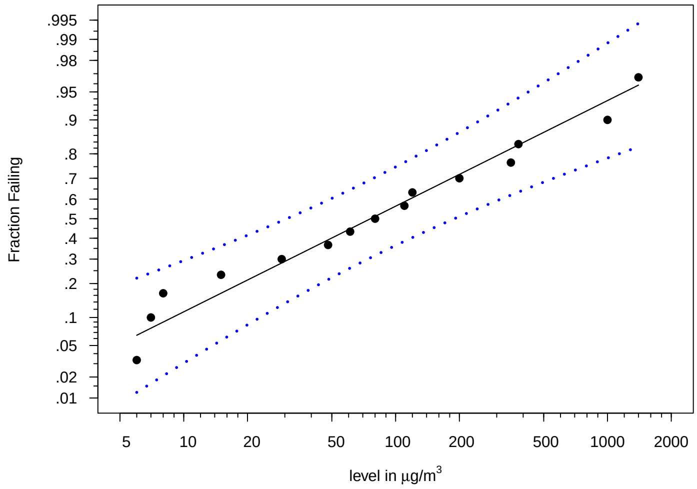
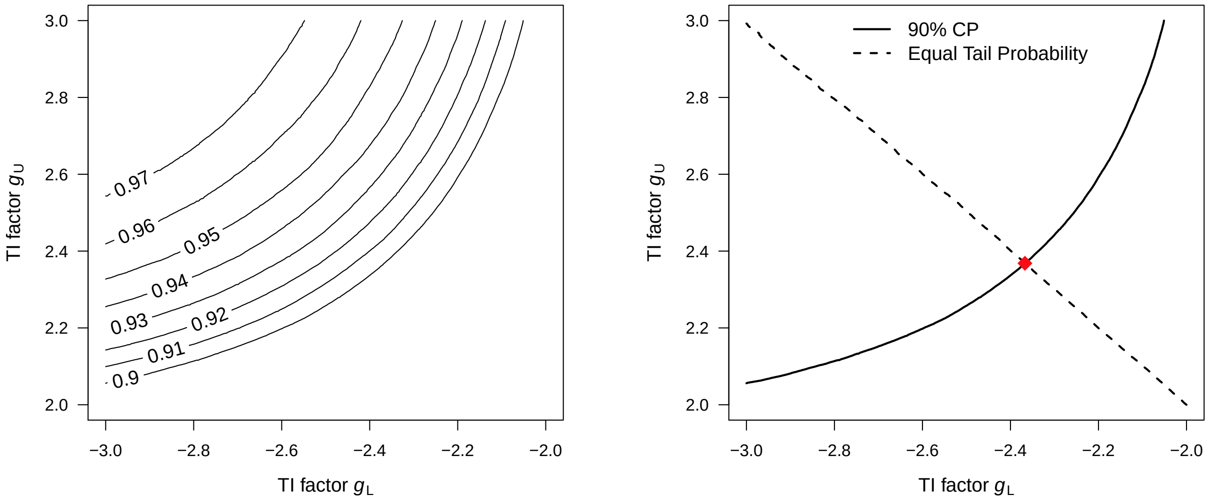


Figure 1: Lognormal probability plot for the air lead level data.

the air lead levels will fall between 1.42 and 4087.48 $\mu g/m^3$. The corresponding control-both-tails TI indicates that there is 90% chance that no more than a proportion 0.1 of the air lead levels will exceed 6118.09 or fall below 0.95 $\mu g/m^3$.

5.2 Pressure Vessel Failure Data

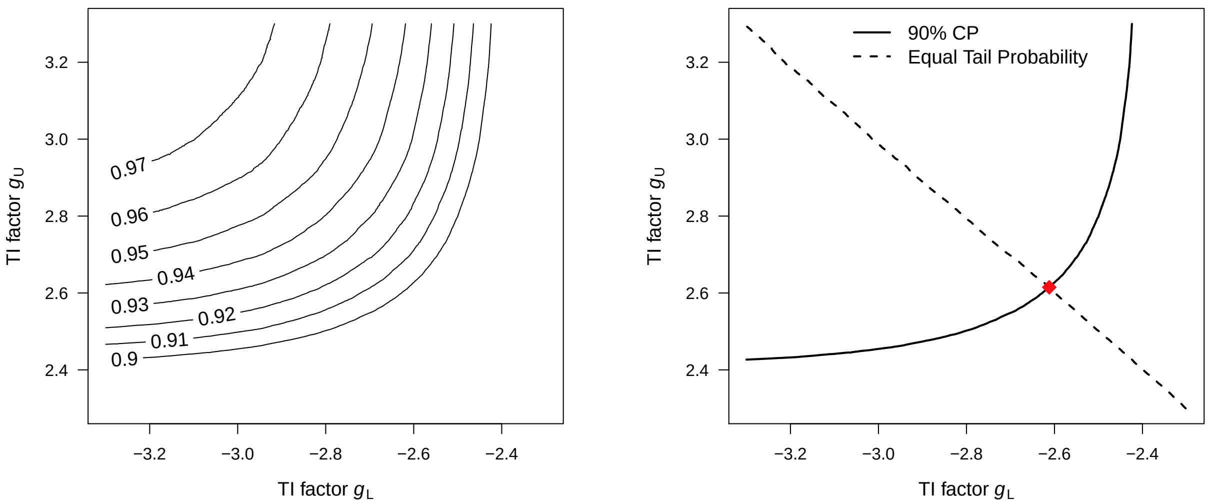
Failure times of 39 pressure vessels were given in Martz and Waller (1982) and used in Howlader and Weiss (1992). The data are shown in Table 3. We treat the largest $n-r = 23$ observations as right censored (i.e., we assume that the 39 pressure vessels were observed until there were $r = 16$ failures). Therefore the data are Type II censored. From Figure 4, we see that both the Weibull and the loglogistic distributions provide good description for the data. TIs under both distributions are computed to characterize the lifetime distribution of the pressure vessels so that one can see the sensitivity with respect to the specification of the distribution. For the content and confidence level ($\beta = 0.9, 1-\alpha = 0.9$), the computed factors and TIs are shown in Table 4.



(a) Factors for TI with $\beta = 0.9$ and different CP

(b) Factors for $(\beta = 0.9, 1 - \alpha = 0.9)$ TI with equal tails

Figure 2: Contour plots of factors for control-the-center TI.



(a) Factors for TI with $\beta = 0.9$ and different CP

(b) Factors for $(\beta = 0.9, 1 - \alpha = 0.9)$ TI with equal tails

Figure 3: Contour plots of factors for control-both-tails TI.

Table 3: Failure times of pressure vessels in hours.

2.2	4.0	4.0	4.6	6.1	6.7	7.9	8.3
8.5	9.1	10.2	12.5	13.3	14.0	14.6	15.0

Table 4: Tolerance intervals for the pressure vessel failure data.

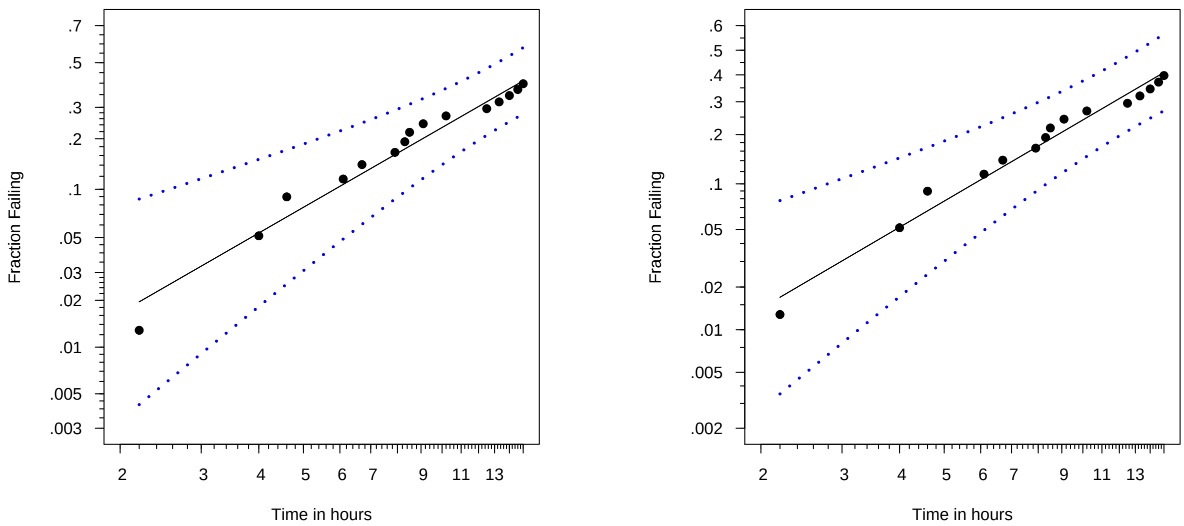
Model	ML Estimates $\widehat{\mu}$	$\widehat{\sigma}$	TI Control-Center	TI Control-Tails
Weibull	3.0796	0.5835	$g_L = -4.09,\ g_U = 2.19$ TI = (2.00, 77.98)	$g'_L = -4.38,\ g'_U = 2.45$ TI = (1.69 90.77)
Loglogistic	2.8979	0.5195	$g_L = -4.06,\ g_U = 4.78$ TI = (2.20, 217.44)	$g'_L = -4.33,\ g'_U = 5.21$ TI = (1.91 272.00)

Based on the TIs using the Weibull distribution, we have 90% confidence that at least a proportion 0.9 of the pressure vessels will have life times between 2 and 77.98 hours, and no more than a proportion 0.1 of the pressure vessels will fail within 1.69 hours or after 90.77 hours. The upper endpoints of the loglogistic distribution TIs are much larger due to extrapolation and the heavier tail of that distribution.

5.3 Locomotive Control Failure Data

Nelson (1982, page 324) gives the miles to failure (in units of 1000 miles) of 96 different locomotive controls. The data are shown in Table 5. Krishnamoorthy and Xie (2011) constructed two-sided TIs using both lognormal and loglogistic distributions to describe the failure time distribution of these locomotive controls. The observation period ended at $x_c =$ 135 thousand miles, by which 37 locomotive controls had failed. Thus the miles to failure of the locomotive controls make a Type I censored dataset. A $(\beta = 0.9, 1-\alpha = 0.9)$ TI is needed to estimate the lifetime and assess the reliability of the controls. Because both lognormal and loglogistic distributions fit the data well, we compare TIs for these two distributions so that one can see the sensitivity with respect to the specification of the distribution. Results of the computed factors and TIs are presented in Table 6. The TIs based on the lognormal distribution are more conservative.

From the calculated TI using the lognormal distribution, we have 90% confidence that at least a proportion 0.9 of these locomotive controls will have life times between 43.67 and 733.08 thousand miles. Also we can be 90% confident that no more than a proportion 0.1 of



(a) Weibull probability plot

(b) Loglogistic probability plot

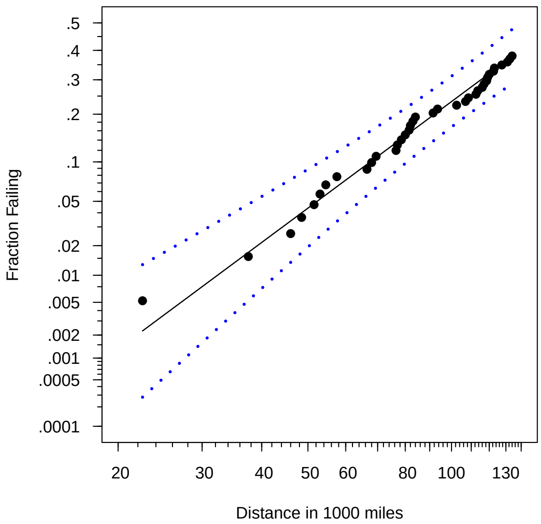
Figure 4: Probability plots for the pressure vessel failure data.

Table 5: Miles to failure of locomotive controls in units of thousands of miles.

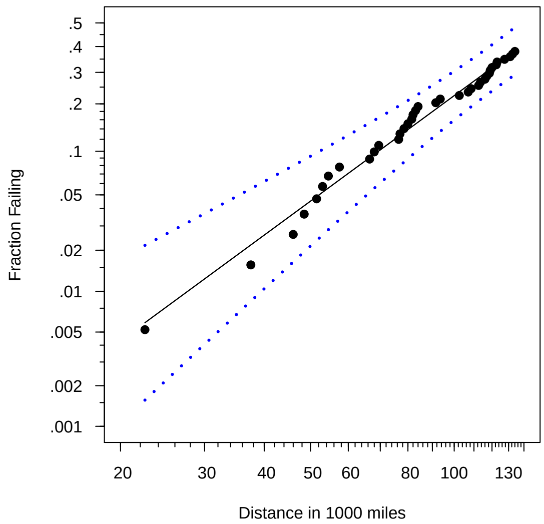
22.5	37.5	46.0	48.5	51.5	53.0	54.5	57.5	66.5	68.0
69.5	76.5	77.0	78.5	80.0	81.5	82.0	83.0	84.0	91.5
93.5	102.5	107.0	108.5	112.5	113.5	116.0	117.0	118.5	119.0
120.0	122.5	123.0	127.5	131.0	132.5	134.0			

Table 6: Tolerance intervals for the locomotive control failure data.

Model	ML Estimates $\widehat{\mu}$	$\widehat{\sigma}$	TI Control-Center	TI Control-Tails
Lognormal	5.1169	0.7055	$g_L = -1.90,\ g_U = 2.10$ TI = (43.67, 733.08)	$g'_L = -1.99,\ g'_U = 2.23$ TI = (41.05 804.38)
Loglogistic	5.0830	0.3837	$g_L = -3.50,\ g_U = 3.78$ TI = (42.02, 687.72)	$g'_L = -3.65,\ g'_U = 3.98$ TI = (39.72 743.84)



(a) Lognormal probability plot



(b) Loglogistic probability plot

Figure 5: Probability plots for the locomotive control failure data.

these locomotive controls will run less than 41.05 or more than 804.38 thousand miles.

6 Simulations for Type I Censoring

Because the TIs with Type I censored data are not exact, this section presents the results of a simulation to evaluate the actual CP. The actual CP depends on the unknown parameters through the expected number of uncensored observations. As the expected number of uncensored observations increases to infinity, the ML estimates approach the true parameter values and the actual CP will approach the nominal confidence level. Thus, to evaluate finite sample properties of the TI procedure, we examine the actual CP of TIs as a function of the expected number of uncensored observations.

We consider both control-the-center and control-both-tails TI for the Weibull distribution with ($\mu = 0, \sigma = 1$). Denote the expected number and expected proportion of uncensored observations as k_f and p_f, respectively. Then, the procedure of the simulation study for the control-the-center TI with equal tails is as follows:

1. Generate $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ from the standard SEV distribution with pre-specified censoring point x_c, which is the p_f lower quantile of the standard SEV distribution. That is $x_c = \log[-\log(1 - p_f)]$. The sample size n is determined by $n = k_f/p_f$. Then calculate ML estimates $(\widehat{\mu}, \widehat{\sigma})$.

2. For simulated data $\boldsymbol{x}$, use **Algorithm 1** to find pairs of (g_L, g_U) with CP equal to the nominal confidence level based on B_2 Monte-Carlo samples. Then identify the unique pair of factors (g_{L0}, g_{U0}) for equal-tail TI by solving (17). Details are given in Section 4.1.

3. To calculate the CP, repeat steps 1 to 2 B_1 times. Denote the simulated data by $\boldsymbol{x}_j^*$, the corresponding ML estimates by $(\widehat{\mu}_j^*, \widehat{\sigma}_j^*)$ and the equal-tail factors by (g_{Lj}, g_{Uj}), $j = 1, \cdots, B_1$. Then calculate TI conditional on each $\boldsymbol{x}_j^*$ as $(\widehat{\mu}_j^* + g_{Lj}\widehat{\sigma}_j^*, \quad \widehat{\mu}_j^* + g_{Uj}\widehat{\sigma}_j^*)$. The CP based on B_1 samples from the standard SEV distribution can be obtained as,

$$\mathrm{CP}(\boldsymbol{\theta}) \approx \frac{1}{B_1}\sum_{j=1}^{B_1} \mathrm{I}\Big[\Phi\left(\widehat{\mu}_j^* + g_{Uj}\widehat{\sigma}_j^*\right) - \Phi\left(\widehat{\mu}_j^* + g_{Lj}\widehat{\sigma}_j^*\right) > \beta\Big].$$

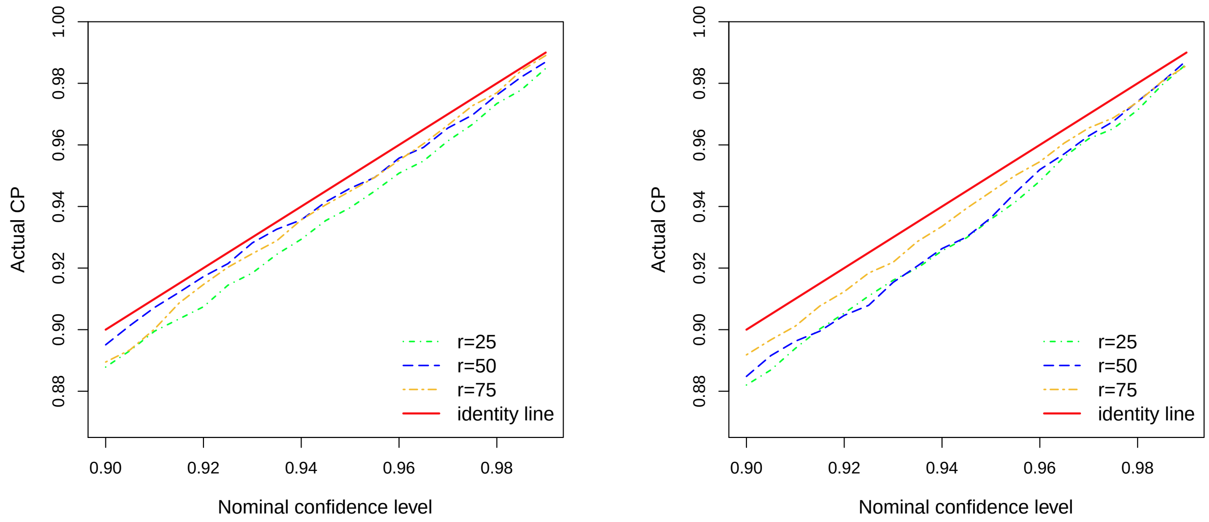
The simulation procedure to study the control-both-tails TI is similar, it only differs in the calculation of the CP.

To see the performance **Algorithm 1** under type I censoring, we consider values of k_f equal to $25, 50, 75$. The expected proportion of uncensored observations p_f is specified as 0.25. Correspondingly, the sample size n takes the values $100, 200, 300$ for both types of TI. The number of type I simulations B_1 is 4000.

The CP compared to the nominal confidence levels are summarized in Figure 6. The lines of control-the-center and control-both-tails TI with $k_f = 75$ are close to the identity line. The actual CP is slightly off the nominal confidence level. Also the CP is closer to the nominal one when the nominal confidence level is larger for all of the values of the expected number of uncensored observations considered. In addition, as the expected number of uncensored observations increases, the CP tends to approach the nominal confidence level for both types of TI. Overall, Figure 6 displays satisfactory results when k_f is moderate to large. The plots on the two panels also show that the CP of control-both-tails TI tends to be lower than those of control-the-center TI when they have the same k_f and nominal confidence level.

7 Conclusions and Areas for Future Research

In this study, we developed a general method to compute control-the-center TIs and control-both-tails TIs for the (log)-location-scale family of distributions. Previous published work did not provide a procedure to calculate two-sided TIs for Weibull (SEV) distribution. When data are complete or Type II censored, our method provides exact factors. For Type I censoring, our method provides factors that give approximate TIs. In this case, the factors are asymptotically accurate. Our simulation study showed that under Type I censored data,



(a) Control-the-center TI with equal tails (b) Control-both-tails TI with equal tails

Figure 6: Plots of the CP vs nominal confidence level for TIs with content 0.9, under Type I censored data.

the CP is close to the nominal confidence level when the expected number of uncensored observations is moderate to large.

The selection of an appropriate lifetime distribution for the data is an important step. Knowledge of the data-generation mechanism can be useful for some applications. If the failure is due to fracture from fatigue in ductile materials, the lognormal distribution can be an appropriate model (see, for example, Chapter 11 of Meeker and Escobar 1998). If the failure is due to fracture from fatigue in brittle materials, the Weibull distribution usually provides a good model. In other situations, probability plots are widely used to select an appropriate distribution.

For the approximate TIs, the CP approaches its nominal confidence level as the expected number of uncensored observations increases. The CP might not be satisfactory when the expected number of uncensored observations is small. It might be possible to investigate alternative methods like bias correction or other methods with another layer of simulation that might provide improved CP.

Sometimes it is also of interest to find the shortest TI. For this purpose, we only need to find the factors on the curve $\mathrm{CP}(g_L, g_U) = 1 - \alpha$ with the absolute difference between g_U and g_L reaching the minimum. One could also develop approaches to compute TIs for distributions outside the (log)-location-scale family, for applications involving regression models, and tolerance regions for multivariate distributions that are often used to characterize the multidimensional output of a process.

Acknowledgments

The authors thank the editor, an associate editor, and two referees, for their valuable comments that helped in improving this paper. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources.

A Proof of Result 1

When a TI is to be computed for a univariate continuous distribution, such as the normal distribution or the Weibull distribution, the CP can be expressed as

$$\mathrm{CP}(\boldsymbol{\theta}) = \int \mathrm{I}\left\{F[\widetilde{T}(\boldsymbol{x},\beta,1-\alpha);\boldsymbol{\theta}] - F[\underset{\sim}{T}(\boldsymbol{x},\beta,1-\alpha);\boldsymbol{\theta}] > \beta\right\} f(\boldsymbol{x};\boldsymbol{\theta})d\boldsymbol{x},$$

where $f(\boldsymbol{x};\boldsymbol{\theta})$ is the joint density function of the data $\boldsymbol{X}$ and the integration is over the region of $\boldsymbol{x}$ values for which $f(\boldsymbol{x};\boldsymbol{\theta}) > 0$. When the TI endpoints can be expressed as functions of model parameter estimates $\widehat{\boldsymbol{\theta}}$, the CP can be computed as

$$\mathrm{CP}(\boldsymbol{\theta}) = \int \mathrm{I}\left\{F[\widetilde{T}(\widehat{\boldsymbol{\theta}},\beta,1-\alpha);\boldsymbol{\theta}] - F[\underset{\sim}{T}(\widehat{\boldsymbol{\theta}},\beta,1-\alpha);\boldsymbol{\theta}] > \beta\right\} h(\widehat{\boldsymbol{\theta}};\boldsymbol{\theta})d\widehat{\boldsymbol{\theta}},$$

where $h(\widehat{\boldsymbol{\theta}};\boldsymbol{\theta})$ is the joint density function (sampling distribution) of the parameter estimator $\widehat{\boldsymbol{\theta}}$ and the integration is over the region of $\widehat{\boldsymbol{\theta}}$ values for which $h(\widehat{\boldsymbol{\theta}};\boldsymbol{\theta}) > 0$.

Following from (7), the CP of a two-sided TI to control the center for the location-scale family of distributions can be expressed as

$$\begin{aligned}\mathrm{CP}(\boldsymbol{\theta}) &= \mathrm{E}_{\widehat{\mu},\widehat{\sigma}}\left(\mathrm{I}\left\{\Phi\left[\frac{\widetilde{T}(\widehat{\mu},\widehat{\sigma},\beta,1-\alpha)-\mu}{\sigma}\right] - \Phi\left[\frac{\underset{\sim}{T}(\widehat{\mu},\widehat{\sigma},\beta,1-\alpha)-\mu}{\sigma}\right] > \beta\right\}\right)\\ &= \mathrm{E}_{\widehat{\mu},\widehat{\sigma}}\left\{\mathrm{I}\left[\Phi\left(\frac{\widehat{\mu}+g_{U(1-\alpha;\beta,n)}\widehat{\sigma}-\mu}{\sigma}\right) - \Phi\left(\frac{\widehat{\mu}+g_{L(1-\alpha;\beta,n)}\widehat{\sigma}-\mu}{\sigma}\right) > \beta\right]\right\}, \qquad (19)\end{aligned}$$

where $\Phi(z)$ is the cdf of a particular standard location-scale distribution and the expectation is with respect to the joint distribution of the estimators $(\widehat{\mu},\widehat{\sigma})$. Following (19),

$$\begin{aligned}\mathrm{CP}(\boldsymbol{\theta}) &= \mathrm{E}_{\widehat{\mu},\widehat{\sigma}}\left\{\mathrm{I}\left[\Phi\left(\frac{\widehat{\mu}-\mu}{\sigma}+g_{U(1-\alpha;\beta,n)}\frac{\widehat{\sigma}}{\sigma}\right) - \Phi\left(\frac{\widehat{\mu}-\mu}{\sigma}+g_{L(1-\alpha;\beta,n)}\frac{\widehat{\sigma}}{\sigma}\right) > \beta\right]\right\}\\ &= \mathrm{E}_{Z_1,Z_2}\left\{\mathrm{I}\left[\Phi\left(Z_1+g_{U(1-\alpha;\beta,n)}Z_2\right) - \Phi\left(Z_1+g_{L(1-\alpha;\beta,n)}Z_2\right) > \beta\right]\right\},\end{aligned}$$

where the expectation is with respect to the joint distribution of $Z_1 = (\widehat{\mu}-\mu)/\sigma$ and $Z_2 = \widehat{\sigma}/\sigma$.

B Proof of Result 2

From Lawless (2003, pp. 562–563), the quantities Z_1 and Z_2 for a location-scale family of distribution are pivotal because the number of uncensored observations r is fixed from sample to sample. When data are Type I censored, the exact pivotal properties do not hold. It can be proved, however, that when the expected proportion of uncensored observations p_f is fixed, Z_1 and Z_2 are asymptotically pivotal.

Let $F(x)$ be the cdf of the population and $F_n(x)$ denote its empirical distribution function. By the Glivenko-Cantelli Theorem,

$$\sup_{x\in\mathbb{R}}|F_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0. \tag{20}$$

That is, the actual proportion of uncensored observations

$$\hat{p}_f = \sum_{i=1}^{n}[I(X_i \leq x_c)]/n \xrightarrow{\text{a.s.}} p_f = F(x_c), \text{for all } x_c.$$

Thus, the actual number of uncensored observations $r = n\hat{p}_f$ is approximately equal to np_f, which is a fixed number.

In particular, for $\boldsymbol{x} = (x_1, \cdots, x_n)$ independently and identically distributed (i.i.d.) as $F(x;\mu,\sigma)$ and $\boldsymbol{x'} = (x'_1, \cdots, x'_n)$ i.i.d. as $F(x;\mu',\sigma')$ with $x'_i = ax''_i + b$, $\mu' = a\mu + b$ and $\sigma' = a\sigma$ (i.e., F is a member of the location-scale family of distributions), when they have the same expected proportion of uncensored observations p_f,

$$\begin{aligned}
L_{\boldsymbol{x}}(\mu,\sigma) =& \frac{1}{\sigma^r}\left[\prod_{i=1}^{r_1}\phi\left(\frac{x_i-\mu}{\sigma}\right)\right]\left[1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right]^{n-r_1}, \\
L_{\boldsymbol{x'}}(\mu',\sigma') =& \frac{1}{(a\sigma)^r}\left[\prod_{i=1}^{r_2}\phi\left(\frac{x'_i-\mu'}{\sigma'}\right)\right]\left[1-\Phi\left(\frac{ax_c+b-\mu'}{\sigma'}\right)\right]^{n-r_2} \\
=& \frac{1}{(a\sigma)^r}\left[\prod_{i=1}^{r_2}\phi\left(\frac{x''_i-\mu}{\sigma}\right)\right]\left[1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right]^{n-r_2} \\
=& \frac{1}{(a\sigma)^r}\left[\prod_{i=1}^{r_1}\phi\left(\frac{x''_i-\mu}{\sigma}\right)\right]\left[1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right]^{n-r_1}\left[\prod_{i=r_1+1}^{r_1+\delta}\phi\left(\frac{x''_i-\mu}{\sigma}\right)\right] \\
& \Big/\left[1-\Phi\left(\frac{x_c-\mu}{\sigma}\right)\right]^{\delta},
\end{aligned}$$

where $r_1 = \sum_{i=1}^{n} I(x_i \leq x_c)$, $r_2 = \sum_{i=1}^{n} I(x'_i \leq ax_c + b)$ are the number of exact observations in $\boldsymbol{x}$ and $\boldsymbol{x'}$, respectively. Set $\mid r_2 - r_1 \mid = \delta$, and ϕ, Φ to be the pdf and cdf of the corresponding standard distribution.

Based on (20), as $n \to \infty$, for a fixed p_f, we have

$$\sup_{x_c \in \mathbb{R}} |r_1/n - p_f| \xrightarrow{\text{a.s.}} 0, \sup_{x_c \in \mathbb{R}} |r_2/n - p_f| \xrightarrow{\text{a.s.}} 0. \text{ Therefore, } \sup_{x_c \in \mathbb{R}} \delta/n \xrightarrow{\text{a.s.}} 0.$$

Hence,

$$L_{\boldsymbol{x}'}(a\mu + b, a\sigma) \xrightarrow{\text{a.s.}} \frac{1}{a^r} \left\{ \frac{1}{\sigma^r} \left[\prod_{i=1}^{r_1} \phi\left(\frac{x_i'' - \mu}{\sigma} \right) \right] \left[1 - \Phi\left(\frac{x_c - \mu}{\sigma} \right) \right]^{n-r_1} \right\} \text{ for all } x_c, \quad (21)$$

where $\boldsymbol{x''} = (x_1'', \cdots, x_n'')$ is an independent sample from the same distribution as $\boldsymbol{x}$. Based on (21), $\hat{\mu}' \approx a\hat{\mu} + b$ and $\hat{\sigma}' \approx a\hat{\sigma}$ (i.e., ML estimators $\hat{\mu}$ and $\hat{\sigma}$ from Type I censored data are approximately equivariant) when n is large enough. Therefore, quantities Z_1 and Z_2 in **Result 2** are approximately pivotal under Type I censored samples of large size.

References

Bergquist, M. L. (2006). *Caution Using Bootstrap Tolerance Limits with Application to Dissolution Specification Limits*. Ph. D. thesis, North Carolina State University.

Eberhardt, K. R., R. W. Mee, and C. P. Reeve (1989). Computing factors for exact two-sided tolerance limits for a normal distribution. *Communications in Statistics - Simulation and Computation 18*, 397–413.

Hahn, G. J. and W. Q. Meeker (1991). *Statistical Intervals - A Guide for Practitioners*. New York: John Wiley & Sons, Inc.

Howe, W. G. (1969). Two-sided tolerance limits for normal populations - some improvements. *Journal of the American Statistical Association 64*, 610–620.

Howlader, H. A. and G. Weiss (1992). Log-logistic survival estimation based on failure-censored data. *Journal of Applied Statistics 19*, 231240.

Krishnamoorthy, K. and T. Mathew (2009). *Statistical Tolerance Regions - Theory, Applications, and Computation*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Krishnamoorthy, K., T. Mathew, and G. Ramachandran (2006). Generalized p-values and confidence intervals: a novel approach for analyzing lognormally distributed exposure data. *Journal of Occupational and Environmental Hygiene 3*, 642–650.

Krishnamoorthy, K. and F. Xie (2011). Tolerance intervals for symmetric location-scale families based on uncensored or censored samples. *Journal of Statistical Planning and Inference 141*, 1170–1182.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data.* John Wiley & Sons.

Martz, H. F. and R. A. Waller (1982). *Bayesian Reliability Analysis.* New York: John Wiley & Sons, Inc.

Meeker, W. Q. and L. A. Escobar (1998). *Statistical Methods for Reliability Data.* John Wiley & Sons.

Nelson, W. (1982). *Applied Life Data Analysis.* New York: John Wiley& Sons, Inc.

Odeh, R. E. and D. B. Owen (1980). *Tables for Normal Tolerance Limits, Sampling Plans, and Screening.* New York: Marcel Dekker, Inc.

Shyu, J. and D. B. Owen (1986a). One-sided tolerance intervals for the two-parameter double exponential distribution. *Communications in Statistics - Simulation and Computation 15*, 101–119.

Shyu, J. and D. B. Owen (1986b). Two-sided tolerance intervals for the two-parameter double exponential distribution. *Communications in Statistics - Simulation and Computation 15*, 479–495.

Weissberg, A. and G. H. Beatty (1960). Tables of tolerance-limit factors for normal distributions. *Technometrics*, 483–500.

Xie, Y., Y. Hong, W. Q. Meeker, and L. A. Escobar (2014). Simultaneous prediction intervals for members of the (log)-location-scale family of distributions. *Statistics Preprints*, Paper 128, http://lib.dr.iastate.edu/stat_las_preprints/128.