Comparisons of Approximate Confidence Interval Procedures for Type I Censored Data

Shuen-Lin JENG
Department of Applied Statistics
Ming-Chuan University
Kuei-Shan, Taoyuan 333
Taiwan
Republic of China
(sljeng@mcu.edu.tw)

William Q. MEEKER
Department of Statistics
Iowa State University
Ames, IA 50011
(wqmeeker@iastate.edu)

This article compares different procedures to compute confidence intervals for parameters and quantiles of the Weibull, lognormal, and similar log-location-scale distributions from Type I censored data that typically arise from life-test experiments. The procedures can be classified into three groups. The first group contains procedures based on the commonly used normal approximation for the distribution of studentized (possibly after a transformation) maximum likelihood estimators. The second group contains procedures based on the likelihood ratio statistic and its modifications. The procedures in the third group use a parametric bootstrap approach, including the use of bootstrap-type simulation, to calibrate the procedures in the first two groups. The procedures in all three groups are justified on the basis of large-sample asymptotic theory. We use Monte Carlo simulation to investigate the finite-sample properties of these procedures. Details are reported for the Weibull distribution. Our results show, as predicted by asymptotic theory, that the coverage probabilities of one-sided confidence bounds calculated from procedures in the first and second groups are further away from nominal than those of two-sided confidence intervals. The commonly used normal-approximation procedures are crude unless the expected number of failures is large (more than 50 or 100). The likelihood ratio procedures work much better and provide adequate procedures down to 30 or 20 failures. By using bootstrap procedures with caution, the coverage probability is close to nominal when the expected number of failures is as small as 15 to 10 or less, depending on the particular situation. Exceptional cases, caused by discreteness from Type I censoring, are noted.

KEY WORDS: Bartlett correction; Bias-corrected accelerated bootstrap; Bootstrap-t; Life data; Likelihood ratio; Maximum likelihood; Parametric bootstrap; Type I censoring.

1. INTRODUCTION

1.1 Objectives

Due to time constraints in life testing, Type I (time) censored data commonly arise from life tests. To make inference on parameters and quantiles of the life distribution, accurate confidence intervals (CI's) are needed. For Type II (failure) censored data (or uncensored data) from location-scale distributions (or log-location-scale distributions), Lawless (1982, p. 147) described pivotal quantities that can be used to obtain exact CI’s for distribution parameters and quantiles (pivotal quantities are functions of the data that have distributions with no unknown parameters and can be inverted to obtain a confidence statement for an unknown parameter). For Type I censoring (more common in practice), however, neither pivotal methods nor other exact CI procedures in general exist.

Today, normal-approximation intervals are used most commonly in commercial software. These procedures, however, may not have coverage probabilities close to nominal values for a small to moderate number of failures, especially for one-sided confidence bounds (CB’s). We evaluate CI procedures to find those that have high accuracy for both one-sided CB’s and two-sided CI’s for situations with heavy censoring and small samples.

We describe some special effects of Type I censoring. With Type I censoring, unlike the complete data or Type II censoring case, the joint distribution of the maximum likelihood (ML) estimators has a discrete component related to the random number of failures. Moreover, t-like quantities have distributions that depend on the pf, the proportion failing at the censoring time. We show that, for these reasons, some bootstrap procedures behave poorly in constructing CI’s for the p quantile when p is close to pf and the expected number of failures is small.

1.2 Related Work

CI’s based on normal-approximation theory for studentized ML estimators (NORM procedure) are easy to calculate and have been implemented in most commercial software packages. Proper transformation of the ML estimator (e.g., the TNORM procedure suggested by Nelson 1982, pp. 330–333) can improve the approximation to the normal distribution. For example, statistics transformed to have a range over a whole real line may provide studentized (or
For inference on the scale parameter, the coverage probabilities for Type I censoring are based on log-likelihood ratio (LLR) tests. For two-parameter exponential samples with Type I censoring, Vander Wie1 and Meeker (1990) showed that, for Type I censored Weibull data from accelerated life tests, LLR-based CI's have coverage probabilities closer to nominal than those from the TNORM procedure.

Doganaksoy and Schmee (1993a) compared four procedures for Type I censored data from Weibull and lognormal distributions. They are NORM, LLR, the standardized LLR, and the LLR with Bartlett correction (LLRBART). They found that LLR-based procedures perform much better than NORM intervals. With complete or moderately censored data, the standardized LLR considerably improves the approximation, especially for small samples (down to 10 expected failures.) Doganaksoy (1995) reviewed likelihood ratio (LR) CI's for reliability and life data analysis applications. He noted that the LLRBART CI's have been used very little in applications due to the computational difficulties of implementation.

Recent research indicates that the bootstrap is a very powerful procedure for computing accurate approximate CI's. The parametric bootstrap procedure approximates the distribution of statistics by simulation or resampling. Hall (1987, 1992), Efron and Tibshirani (1993), and Shao and Tu (1995) described bootstrap theory and methods in detail.

Robinson (1983) applied a bootstrap procedure to location-scale distributions. The statistics used for constructing CI's are pivotal quantities in the case of complete or Type II censored data. He used the method to find CI's for multiple time-censored progressive data and used simulation to evaluate coverage probabilities.

The parametric bootstrap-t (PBT) procedure is second-order accurate under smoothness conditions (Efron 1982). The percentile procedure (Efron 1981) is very easy to implement but usually is only first-order accurate for one-sided CB's. The bias-corrected procedure (BC, Efron 1982) generally has better performance than the percentile procedure. The bias-corrected accelerated procedure (BCA, Efron 1987) provides an alternative, more accurate, procedure to construct CI's that usually improves the performance of the percentile and BC procedures in complete samples.

The signed-root LLR (SRLLR) statistic has an approximate standard normal distribution in large samples (Barndorff-Nielsen and Cox 1994, p. 101). The modified SRLLR procedure (Barndorff-Nielsen 1986, 1991) is third-order accurate in complete samples, but much more effort is needed to get the modification term. Using bootstrap simulation to obtain the sampling distribution of the SRLLR statistic (PBSRLLR), instead of using the large-sample approximate normal distribution, improves the procedure's coverage probabilities, especially for one-sided CB's. To construct CI's that have approximately equal lower and upper error probabilities, one can combine lower and upper one-sided CB's based on the PBSRLLR procedure. This approach is much better than simply using simulation to approximate the distribution of the positive LLR statistic.

1.3 Overview

The remainder of this article is organized as follows. Section 2 describes the model and the estimation method. Section 3 provides details of the procedures for finding approximate CI's for Weibull, lognormal, and other log-location-scale distributions. Section 4 describes the design of the simulation experiment. Section 5 presents the general results from the simulation experiment. Section 6 contains conclusions from the experiment and suggestions for use in applications. Section 7 discusses some special effects of Type I censoring that lead to poor performance of some simulation-based CI/CB procedures. Discussion and some directions for future research are given in Section 8.

2. MODEL AND ESTIMATION

Extensive evaluations of the properties of CI/CB procedures were done for the Weibull distribution, with less extensive evaluations for the lognormal distribution. We, however, only describe the details for the Weibull distribution. Results for the lognormal were similar, and there is little doubt that similar results would also be obtained with other log-location-scale distributions.

2.1 Model

If $T$ has a Weibull distribution, then $Y = \log(T)$ has a smallest extreme value (SEV) distribution with density

$$f_Y(y) = \frac{1}{\sigma} \exp \left( \frac{y - \mu}{\sigma} - \exp \left( \frac{y - \mu}{\sigma} \right) \right),$$

and cdf

$$F_Y(y) = 1 - \exp \left( - \exp \left( \frac{y - \mu}{\sigma} \right) \right),$$

where $\mu$ and $\sigma$ are location and scale parameters of the distribution of $Y$. The $p$ quantile of the SEV distribution is $y_p = F_Y^{-1}(p) = \mu + c_p \sigma$, where $c_p = \log[1 - \log(1 - p)]$ is the $p$ quantile of the standardized $(\mu = 0, \sigma = 1)$ SEV distribution. The traditional Weibull scale and shape parameters are $\alpha = \exp(\mu)$ and $\beta = 1/\sigma$, respectively.

2.2 ML Estimation

We use $\hat{\mu}$ and $\hat{\sigma}$ to denote the ML estimators of the SEV parameters. Because of the invariance property of ML estimators, $\hat{y}_p = \hat{\mu} + \hat{c}_p \hat{\sigma}$ is the ML estimator of the $p$ quantile of the SEV distribution. Moreover, the ML estimators of the Weibull parameters are $\hat{\alpha} = \exp(\hat{\mu})$ and $\hat{\beta} = 1/\hat{\sigma}$. The ML estimator of the Weibull $p$ quantile is $\hat{y}_p = \exp(\hat{y}_p)$. More generally, the ML estimator of a function $g(\mu, \sigma)$ is $\hat{g} = g(\hat{\mu}, \hat{\sigma})$. For any particular function of interest, it is possible to reparameterize by defining a one-to-one transformation, $g(\mu, \sigma) = (g_1(\mu, \sigma), g_2(\mu, \sigma)) = \theta$, that contains...
the function of interest among its elements. For example, $g_1(\mu, \sigma)$ could be a distribution quantile or failure probability. Then ML fitting can be carried out for this new parameterization in a manner that is the same as that described previously for $(\mu, \sigma)$. This provides a procedure for obtaining ML estimates and likelihood CI's for any scalar or vector function of $(\mu, \sigma)$. For more details see Lawless (1982, chap. 4) or Meeker and Escobar (1998, sec. 8.3.3).

Let $\theta = (\theta_1, \theta_2)$ be the unknown parameter vector, where $\theta_1$ is the parameter of interest and $\theta_2$ is a nuisance parameter. Typically $\theta$ could be $(\mu, \sigma)$ or $(\exp, \sigma)$. We use $L(\theta)$ to denote the likelihood and $t_c$ to denote the specified censoring time. Let $t_1, \ldots, t_n$ be $n$ observations (e.g., failure or censoring times) from a life test. If the observations are independent, then the censored data likelihood is

$$L(\theta) = \prod_{i=1}^{n} f_Y(\log(t_i); \theta)^{\delta_i} [1 - F_Y(\log(t_c); \theta)]^{1-\delta_i},$$

where $\delta_i = 1$ if $t_i$ is a failure time and $\delta_i = 0$ if observation $i$ is censored at $t_c$.

### 3. CONFIDENCE INTERVAL/BOUND PROCEDURES

This section describes the different CI/CB procedures that are evaluated in this article. Table 1 shows the abbreviations for each procedure. Let $C_{n,1-\alpha}$ denote an approximate CI for $\theta_1$ with nominal coverage probability $1 - \alpha$, where $n$ is the sample size. The procedure for obtaining $C_{n,1-\alpha}$ is said to be $k$th-order accurate if $P_1(\theta_1 \in C_{n,1-\alpha}) = 1 - \alpha + O(n^{-k/2})$. If there is no $O(\cdot)$ term in the equation, we say that the procedure for $C_{n,1-\alpha}$ is “exact.” The following subsections show how to compute an approximate two-sided $100(1-\alpha)$% CI for each CI procedure used in the comparison. One-sided CB’s are obtained by using the appropriate endpoint of a two-sided CI, with a corresponding adjustment to the confidence level.

#### 3.1 Normal-Approximation Procedures

**Normal-Approximation Procedure (NORM).** Suppose that $\hat{\theta}$ is the ML estimator of the parameter vector $\theta$. Under the usual regularity conditions, $\hat{\theta}$ is asymptotically normal and efficient (Serfling 1980, p. 148). Let $I_0$ denote the Fisher information matrix, and let $n[I_0^{-1}]^{1/2}$ be an estimator that converges to $[g'(\theta_1)]^{-1}$ in probability. Then the distribution of $(\hat{\theta}_1 - \theta_1)/\sqrt{n}[g'(\theta_1)]^{-1}$ is approximately $N(0, 1)$ in large samples. Thus a normal-approximation $100(1-\alpha)$% CI can be obtained from $\hat{\theta}_1 \pm z_{(1-\alpha)/2}\sqrt{n}[g'(\theta_1)]^{-1}$, where $z_{(1-\alpha)/2}$ is the N(0, 1) distribution 1 - $\alpha/2$ quantile. In this article, $n[I_0^{-1}]^{1/2}$ is obtained from the inverse of the local estimate of $I_0$ (e.g., Nelson 1982, p. 377).

**Transformed Normal-Approximation Procedure (TNORM).** When an ML estimator $\hat{\theta}_1$ has its range on only part of the real line, a monotone function $g(\theta_1)$ with continuous derivatives and with range on the entire real line generally has a better normal approximation (Nelson 1982, p. 331). Let $g'(\theta_1)$ denote the first derivative of $g(\theta_1)$, and let $n[I_0^{-1}]^{1/2}$ be an estimator that converges to $[g'(\theta_1)]^{-1/2}$ in probability. The TNORM procedure is based on the normal approximation $[g'(\theta_1) - g(\theta_1)]/\sqrt{n[I_0^{-1}]} \approx N(0, 1)$. Then the TNORM CI procedure uses $g^{-1}[g(\theta_1) \pm z_{(1-\alpha)/2}\sqrt{n[I_0^{-1}]^{1/2}]}$, where $z_{(1-\alpha)/2}$ is the 1 - $\alpha/2$ quantile of the N(0, 1) distribution. Typically $g$ could be the log function for a scale parameter or for positive quantile parameters and the logit or tan$^{-1}$ function for a probability parameter. In this article $n[I_0^{-1}]^{1/2}$ is obtained, using the delta method, as $[g'(\theta_1)]^{-1}I_0^{1/2}$, where $I_0$ is the local estimate of $I_0$.

#### 3.2 Likelihood Ratio Procedures

**Log-Likelihood Ratio Procedure (LLR).** The profile likelihood for $\theta_1$ is defined as

$$R(\theta_1) = \max_{\theta_2} \left[ L(\theta_1, \theta_2) \right],$$

Let $W = W(\theta_1) = -2\log R(\theta_1)$. From Serfling (1980, sec. 4.4), the limiting distribution of $W$ is $\chi^2_1$. Let $\chi^2_{(1-\alpha)}$ denote the 1 - $\alpha$ quantile of the $\chi^2$ distribution with 1 df. The equation $W(\theta_1) - \chi^2_{(1-\alpha)} = 0$ generally has two roots, one less than and one greater than $\theta$. The LLR CI procedure uses these roots as the lower and upper confidence bounds, respectively.

**LLR Bartlett-Corrected Procedure (LLRBART).** Before the expectation of $E(W)$ is equal to the mean of the $\chi^2$ distribution, the distribution of $W/E(W)$ will be better approximated by the $\chi^2_1$ distribution (Bartlett 1937). In general one must substitute an estimate for $E(W)$ computed from one’s data. For complicated problems (e.g., those involving censoring) it is necessary to estimate $E(W)$ by using simulation, as described by Doganaksoy and Schmee (1993a). Then, similar to the LLR procedure, the LLRBART CI procedure uses the two roots of $W(\theta_1)/E(W) - \chi^2_{(1-\alpha)} = 0$ as the lower and upper confidence bounds, respectively.

#### 3.3 Parametric Bootstrap Procedures

The following procedures use the “bootstrap principle” or Monte Carlo evaluation of sampling distributions. Suppose that a statistic $S$ is a function of random variables with a distribution that depends on the parameter $\theta$. The parametric bootstrap version $S^*$ of $S$ is the same function but based on data (“bootstrap sample”) simulated using $\theta$ in place of the unknown $\theta$. The distribution of $S^*$ is eas-

**Table 1. Abbreviations for CI/CB Procedures**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORM</td>
<td>Normal-approximation</td>
</tr>
<tr>
<td>TNORM</td>
<td>Transformed normal-approximation</td>
</tr>
<tr>
<td>LLR</td>
<td>Log-likelihood ratio</td>
</tr>
<tr>
<td>LLRBART</td>
<td>Log-likelihood ratio Bartlett corrected</td>
</tr>
<tr>
<td>PBPT</td>
<td>Parametric bootstrap-t</td>
</tr>
<tr>
<td>PBT</td>
<td>Parametric bootstrap-t</td>
</tr>
<tr>
<td>PD39LLR</td>
<td>Parametric bootstrap signed square root LLR</td>
</tr>
<tr>
<td>PBP</td>
<td>Parametric bootstrap percentile</td>
</tr>
<tr>
<td>PBBCA</td>
<td>Parametric bootstrap bias-corrected accelerated</td>
</tr>
<tr>
<td>PBBC</td>
<td>Parametric bootstrap bias-corrected</td>
</tr>
</tbody>
</table>
ily obtained by simulation. Throughout this article, we used 2,000 bootstrap samples to obtain approximate distributions of \( S^* \).

**Parametric Bootstrap t Procedure (PBT) (Efron 1982)**
Let \( \hat{\theta}_1 \) be the ML estimator of \( \theta_1 \) and let \( \hat{\theta}_1^* \) be the ML estimator from a bootstrap sample. Moreover, let \( z_{g(\hat{\theta}_1)} \) be the \( \alpha \) quantile of the distribution of \( Z_{g(\hat{\theta}_1)} = \frac{(\hat{\theta}_1 - \theta_1)}{\hat{\sigma}(\hat{\theta}_1)} \), where \( \hat{\sigma}(\hat{\theta}_1) \) is the bootstrap version of \( \hat{\sigma}(\hat{\theta}_1) \). In this article, we choose \( \hat{\sigma}(\hat{\theta}_1) \) to be the same as in the NORM procedure. The PBT CI procedure uses \( [\hat{\theta}_1 - z_{g(\hat{\theta}_1)} \hat{\sigma}(\hat{\theta}_1), \hat{\theta}_1 - z_{g(\hat{\theta}_1)} \hat{\sigma}(\hat{\theta}_1)] \).

We follow their procedure by using parametric bootstrap to obtain the bias-correction term and using jackknife to obtain the acceleration constant.

**Parametric Transformed Bootstrap-t Procedure (PTBT)**
Let \( g \) be a smooth monotone function generally chosen such that \( g(\hat{\theta}_1) \) has range on whole real line. Let \( \hat{\theta}_1 \) be the ML estimator of \( \theta_1 \), and let \( \hat{\theta}_1^* \) be the bootstrap version ML estimator. Let \( z_{g(\hat{\theta}_1)} \) be the \( \alpha \) quantile of the distribution of \( Z_{g(\hat{\theta}_1)} = \frac{(g(\hat{\theta}_1) - g(\hat{\theta}_1))}{\hat{\sigma}(\hat{\theta}_1)} \), where \( \hat{\sigma}(\hat{\theta}_1) \) is the bootstrap version of \( \hat{\sigma}(\hat{\theta}_1) \). In this article, we choose \( \hat{\sigma}(\hat{\theta}_1) \) to be the same as in the TNORM procedure. When \( g \) is monotone increasing, the PTBT CI procedure for \( \theta_1 \) uses \( [g^{-1}(g(\hat{\theta}_1) - z_{g(\hat{\theta}_1)} \hat{\sigma}(\hat{\theta}_1)), g^{-1}(g(\hat{\theta}_1) - z_{g(\hat{\theta}_1)} \hat{\sigma}(\hat{\theta}_1))] \). When \( g \) is monotone decreasing, the order of the endpoints is reversed.

**Parametric Bootstrap Signed Square Root LLR Procedure (PBSR\( LLR \))**
Let \( V(\theta_1) = \text{sign}(\hat{\theta}_1 - \theta_1)[-2 \log L(\theta_1)]^{1/2} \) denote the signed square root of the LLR statistic. In large samples, the distribution of \( V(\theta_1) \) can be approximated by a normal distribution (Barndorff-Nielsen and Cox 1994, p. 101). Approximating by simulation, however, captures the asymmetry of the distribution and hence provides a better approximation for finding CB's for \( \theta_1 \). Suppose that \( v_{\theta_1} \) is the \( \alpha \) quantile of the bootstrap distribution of \( V(\hat{\theta}_1) \). Then, similar to the LLR procedure, the PBSR\( LLR \) CI procedure uses the roots of \( V(\hat{\theta}_1) - v_{\theta_1} = 0 \) and \( V(\hat{\theta}_1) + v_{\theta_1} = 0 \) as the lower and upper confidence bounds, respectively.

**Parametric Bootstrap Percentile Procedure (PBP), Parametric Bootstrap Bias-Corrected Procedure (PBB), and Parametric Bootstrap Bias-Corrected Accelerated Procedure (PBBCA)**

**3.4 Using Bootstrap Simulation With Single and Multiple Censoring**

The simulation-based parametric bootstrap methods described in Section 3.3 are based on sampling from the assumed distribution using Type I censoring at a specified point in time. See section 4.13 of Meeker and Escobar (1998) for a description of computationally efficient methods for generating such censored samples.

In many applications one will encounter multiple censoring (observations censored at different points in time) such censoring arises for several different reasons, including staggered entry of units into a study and multiple failure modes [see sec. 2.3 of Meeker and Escobar (1998) for further discussion of different kinds of censoring mechanisms]. Simulation can still be used in such situations. Based on asymptotic theory, limited existing results in the literature (especially Robinson 1983) and insights provided by our results, we would expect that the general results observed in our study would also apply to these more complicated censoring patterns. Use of pure parametric simulation would, however, require that the underlying censoring mechanism (or its distribution, in the case of random censoring) be specified exactly so that it could be mimicked in the simulation. In some situations, the details of the censoring mechanism may not be known, and it might not be possible to infer these details from the data. Another alternative, for such situations, is to use ideas from resampling. That is, following the nonparametric bootstrap paradigm, bootstrap samples can be selected by sampling with replacement from the available failure and censored observations. As long as the number of distinct censoring and failure times is reasonably large (say more than 10 or so) and the distributions of the failure and censoring times overlap to some degree, the coverage properties of the procedure should be similar to that of the fully parametric sampling method. This is suggested by the resulting approximate continuity of the bootstrap distribution, as indicated in appendix I of Hall (1992).

**3.5 Numerical Examples**

Figure 1 shows a probability plot for the ball-bearing fatigue data (Lawless 1982, p. 228). Table 2 shows numerical values for two-sided approximate 90% CI's (lower and upper one-sided approximate 95% CI's) computed from...
these data after being artificially censored at both 40 million cycles (3 of 23 failing) and 60 million cycles (11 of 23 failing). As expected, the intervals tend to be much wider for cycles (3 of 23 failing) and 60 million cycles (11 of 23 failing). With $r = 0$, ML estimates do not exist. With $r = 1$, the log-likelihood can be poorly behaved and LR intervals of reasonable length may not exist. Therefore, we give results conditionally on the cases with $r > 1$ and report the observed nonzero proportions that resulted in $r < 1$.

### 4.2 Coverage Probability Comparisons

Let $1 - \alpha$ be the nominal coverage probability (CP) of a CI procedure, and let $1 - \tilde{\alpha}$ denote the corresponding Monte Carlo estimate. The standard error of $\tilde{\alpha}$ is approximately $\alpha(1 - \alpha)/n_x^{1/2}$, where $n_x$ is the number of Monte Carlo simulation trials. For one-sided 95% CB's from 2,000 simulations, the standard error of the CP estimate is $[0.05(1 - 0.95)/2,000]^{1/2} = 0.0049$. Thus, the Monte Carlo error is approximately ±1%. We say the procedure is adequate if the CP is within ±2% error for 95% CB and 90% CI procedures.

If a coverage probability is greater than (less than) $1 - \alpha$, then the CI procedure is conservative (anticonservative). We say that coverage probability is approximately symmetric when the CP's of the lower and upper CB's are approximately the same.

### 5. SIMULATION EXPERIMENT RESULTS

This section presents a summary of the most interesting and useful results from the simulation experiment. Table 3 shows the number of Monte Carlo simulations that had only 0 or 1 failure. Those cases were excluded from coverage probability computation. With $E(r) > 10$, there were no Monte Carlo simulations that had fewer than two failures.

#### 5.1 One-sided CB's

Let UCB (LCB) denote an upper (lower) confidence bound. Figure 2 shows the coverage probability of the one-sided approximate 95% CB's for the parameter $r$ from 10 procedures for 5 cases of proportion failing. This figure shows that the TNORM procedure performs considerably better than the others.

### Table 2. Comparison of Confidence Intervals for $\sigma$, $t_1$, and $t_5$

Based on the Ball-Bearing Data Artificially Censored at 40 and 60 Million Cycles

<table>
<thead>
<tr>
<th>Procedure</th>
<th>$\sigma$</th>
<th>$t_1$</th>
<th>$t_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NORM</td>
<td>[.01, .09]</td>
<td>[17.97, 52.73]</td>
<td>[.69, 149.92]</td>
</tr>
<tr>
<td>TNORM</td>
<td>[.14, .28]</td>
<td>[21.62, 57.80]</td>
<td>[.31, 192.95]</td>
</tr>
<tr>
<td>LLR</td>
<td>[.17, .66]</td>
<td>[14.18, 82.04]</td>
<td>[.47, 870.49]</td>
</tr>
<tr>
<td>LLLR</td>
<td>[.16, 1.78]</td>
<td>[17.35, 68.16]</td>
<td>[.47, 1099.20]</td>
</tr>
<tr>
<td>PBT</td>
<td>[.17, 2.09]</td>
<td>[17.22, 53.30]</td>
<td>[.59, 367.14]</td>
</tr>
<tr>
<td>PTBT</td>
<td>[.18, 2.02]</td>
<td>[20.48, 52.48]</td>
<td>[.53, 176.38]</td>
</tr>
<tr>
<td>PSSRLLR</td>
<td>[.18, 2.07]</td>
<td>[19.41, 76.29]</td>
<td>[.52, 2509.63]</td>
</tr>
<tr>
<td>PBSR</td>
<td>[.08, 1.00]</td>
<td>[21.89, 44.33]</td>
<td>[.45, 236.71]</td>
</tr>
<tr>
<td>PBBC</td>
<td>[.13, 1.24]</td>
<td>[22.80, 45.04]</td>
<td>[.49, 650.15]</td>
</tr>
<tr>
<td>PPBCA</td>
<td>[.14, 1.30]</td>
<td>[20.29, 44.21]</td>
<td>[.47, 358.48]</td>
</tr>
</tbody>
</table>

### Table 3. Number of Cases in Which $r = 0$ or 1 in 2,000 Monte Carlo Simulations of the Experiment

<table>
<thead>
<tr>
<th>$E(r)$</th>
<th>.01</th>
<th>.05</th>
<th>.10</th>
<th>.30</th>
<th>.50</th>
<th>.70</th>
<th>.90</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>379(395)</td>
<td>365(383)</td>
<td>376(367)</td>
<td>308(298)</td>
<td>235(218)</td>
<td>160(167)</td>
<td>63(55)</td>
</tr>
<tr>
<td>5</td>
<td>88(79)</td>
<td>72(74)</td>
<td>68(67)</td>
<td>59(52)</td>
<td>23(21)</td>
<td>11(7)</td>
<td>1(0)</td>
</tr>
<tr>
<td>7</td>
<td>17(14)</td>
<td>16(12)</td>
<td>13(10)</td>
<td>3(5)</td>
<td>1(1)</td>
<td>0(0)</td>
<td>1(0)</td>
</tr>
<tr>
<td>10</td>
<td>0(0)</td>
<td>0(0)</td>
<td>0(0)</td>
<td>0(0)</td>
<td>0(0)</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
</tbody>
</table>

NOTE: The expected numbers (rounded to the nearest integer) are shown inside parentheses.
better than the NORM procedure, but even TNORM requires large samples (e.g., larger than 50) before the CP approximation is adequate. The LLR and LLRBART procedures perform better. The PBSRLLR, PBT, and PTBT procedures always provide excellent approximations even for the $E(r) = 3$ case, dominating all of the other procedures evaluated here.

For estimating distribution quantiles, the situation is more complicated. Figure 3 gives CP versus $E(r)$ for CI procedures applied to the Weibull distribution quantile $t_{\alpha}$. As with $\sigma$, the LLR procedure provides a substantial improvement over the NORM and TNORM procedures. LLRBART provides little or no improvement over LLR. Among the other simulation-based procedures, the PBSRLLR procedure provides an excellent approximation in all cases when $E(r) \geq 15$. It also does well for $E(r)$ as small as 3, except when estimating $t_{\alpha}$ when $p \approx p_f$. We refer to this as the “exceptional case.” The bootstrap-t procedures are transformation dependent, but using a reasonable default transformation (e.g., log for a positive parameter), PTBT provides, in other than the exceptional case, good coverage properties at a small fraction of the computational
Figure 3. Coverage Probability Versus Expected Number of Failures Plot of One-Sided Approximate 95% CI's for Parameter $t_f$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$'s (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

With no transformation, the properties are poor, as shown in the PBT results. The PBBC and PBBCA percentile bootstrap procedures, relative to the simple NORM and TNORM procedures, offer useful improvements in coverage probability accuracy for $E(r) > 15$ but do not seem to offer any advantage over the PBSRLLR and PTBT procedures.

TNORM is generally more accurate than NORM for $E(r) > 30$. The approximation of CP is still crude and depends on $p_f$. UCB's (LCB's) are conservative when $p < p_f$ ($p > p_f$) and are anticonservative when $p > p_f$ ($p < p_f$), except that, when $p$ is close to $p_f$, both are conservative. This change as one crosses $p_f$ was also noted in the results of Ostrouchov and Meeker (1988) and Doganaksoy and Schmee (1993a) and will be explored further in the discussion in Section 7.

Figure 4, for $p_f = .1$, gives CP's for bootstrap procedure for $\sigma$ and several quantities for $E(r) = 15$, the point at which some of the bootstrap procedures begin to perform well. This figure shows clearly the potential problems involved with the naive use of the PBT and PBP procedures. The figure also shows that the PTBT and especially the
PBSRLLR procedures work well with some inaccuracy in the PTBT procedure near the exceptional case.

5.2 Two-Sided Cl's

As shown in Section 5.1, CP's tend to be conservative on one side and anticonservative on the other side. With two-sided intervals, there is an averaging effect, and the overall CP approximations tend to be better.

Figure 5 shows the CP of the two-sided 90% CI procedures for the Weibull \( t_1 \) quantiles. Similar plots (not shown here) were made for \( \sigma \) and other quantiles. The LLR procedure has reasonably accurate coverage probabilities, even for \( E(r) \) as small as 15. Unlike the one-sided intervals, for two-sided intervals, LLRBART provides noticeable improvement, down to \( E(r) = 7 \), especially when the proportion failing is greater than .5.

The PBSRLLR and PTBT procedures provide excellent approximations when \( p > p_f \) especially when \( p_f \) is small (< .1). In the exceptional case, however, when \( p_f \) is close to \( p \), both procedures have a CP that is lower than nominal. In this case the PBSRLLR procedure is better than the PTBT procedure and provides an adequate approximation for \( E(r) \geq 15 \). Detailed results on confidence intervals for \( \sigma \) (not shown here) indicate that the LLRBART, PBT, PTBT, and PBSRLLR procedures all provide excellent approximations to the two-sided coverage probabilities. It is important to recognize, however, that in most applications where two-sided intervals are reported, there is important interest in considering separately the effects of being outside on one side or the other.

5.3 Expected Interval Length

Interval length is another criterion for comparing two-sided CI's. With the same coverage probability, procedures that provide shorter intervals are better. Figures showing the mean interval length of the 2,000 two-sided 90% CI's for parameters \( \sigma \) and \( t_1 \) using 10 different procedures for 5 values of \( p_f \) were given by Jeng (1998).

In comparing CI widths (or, more precisely average width), it is preferable to compare intervals with nearly the same CP. Otherwise, procedures with conservative CP's tend to be wider than anticonservative procedures (something that was easy to see in our results). When estimating \( \sigma \), with constant \( E(r) \), the mean interval length decreases slightly as \( p_f \) increases. For quantiles, again with constant \( E(r) \), interval length tends to increase as \( p \) exceeds \( p_f \). This is a result of extrapolation in time, as predicted by asymptotic theory (e.g., figs. 10.5 and 10.6, Meeker and Escobar 1998).

6. OTHER RESULTS, CONCLUSIONS, AND RECOMMENDATIONS

A smaller simulation experiment was conducted for the lognormal distribution. The results for the lognormal distribution are consistent with what we have reported in Section 5. We draw the following conclusions and recommendations for Weibull and lognormal distributions. We expect that these findings will hold in general for log-location-scale distributions.

Normal-approximation CI's (NORM and TNORM), although still commonly used in practice (e.g., in many statistical software packages), may not be adequate when the expected number of failures is less than 50. For the one-sided case, we see that \( E(r) = 100 \) is needed to provide a good approximation to the nominal coverage probability. If a positive parameter is of interest, the usual log transformation, which makes the ML estimator have range over whole real line, is suggested. Doing this assures that the CI endpoints will always lie in the parameter space and usually (but not always) provides a somewhat better coverage probability for any proportion failing.

Our findings for the normal approximation and LR procedures are consistent with results of Ostrouchov and Meeker (1988), Doganaksoy and Schmee (1993a), and Doganaksoy (1995). This article, however, focuses more on the asymmetry of coverage probability for one-sided CI's, as well as cases with heavy censoring and a small expected number of failures.

Some bootstrap procedures provide better coverage probability accuracy. Using the bootstrap-t without a proper transformation, however, may not perform any better than the normal-approximation procedure. It is important to use the bootstrap-t procedure carefully.

The bootstrap percentile procedures are easy to implement, and they improve the normal-approximation proce-
CONFIDENCE INTERVAL PROCEDURES FOR TYPE I CENSORED DATA

Figure 5. Coverage Probability Versus Expected Number of Failures Plot of Two-Sided 90% CI's for Parameter $t_i$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to $p_f$ (.01, .1, .3, .5, 1).

dure in many (but not all) cases. The accuracy of the parametric bootstrap percentile (PBP), bias corrected (PBBC), and bias-corrected accelerated (PBBCA) procedures depend on the expected number of failures, the proportion failing, and the parameters of interest. When the proportion failing is greater than .1, the PBBCA procedure has better performance than the PBBC procedure for quantile parameters. In heavily censored cases ($p_f < .1$), however, the PBBCA procedure is generally worse. This is probably due to difficulty in estimating the acceleration constant under heavy censoring.

The parametric bootstrap-t with transformation (PTBT) and bootstrap signed-root log-likelihood ratio (PBSRLLR) procedures provide more accurate results over all different numbers of failures, proportion failing, and parameters of interest except for the case in which parameter of interest is $t_p$ and $p$ is close to proportion failing $p_f$. Moreover, upper and lower coverage probabilities are approximately equal, which is important when one-sided CB’s are needed or when the cost of being wrong differs importantly from one side to the other of a two-sided interval. Although the PBSRLLR procedure is more accurate in small samples (say
Figure 6. Weibull Distribution Life-Test Simulation. The results show 100 ML estimates \( F(t) \) based on data simulated from a Weibull distribution with parameters \( \mu = 0, \sigma = 1 \), sample size \( n = 50 \), with Type I censoring at a time corresponding to \( p_f = .1 \) so \( E(r) = 5 \).

when \( E(r) < 10 \), the bootstrap-\( t \) with transformation requires much less computational effort than the PBSRLLR procedure.

In general, when the expected number of failures is smaller than 50 (20), the LR-based procedures can be recommended for finding one-sided CB’s (two-sided CI’s). For smaller \( E(r) \), the PTBT and PBSRLLR procedures can be recommended except for the case in which the quantity of interest is \( t_1 \) when \( E(T) \) is to be estimated and \( p \) is close to the proportion failing \( p_f \). Then, as seen in Figure 6 and as we can be shown analytically, the approximation is excellent for small \( r \). This approximation implies that

\[
\hat{\mu} \approx \log(t_c) - \Phi^{-1}(r/n)\hat{\sigma}.
\]
For the PTBT method to work well, the distribution of $Z_{i,n}$ should not depend on any unknown parameters. Figure 8 compares the bootstrap distributions of $Z_{i,n}$ for censoring times corresponding to $p_f = .1$, .2, and .5 and samples sizes $n = 50, 25$, and $10$, respectively, so that $E(r) = 5$ in each case. For each $p_f$, the figure shows bootstrap distributions corresponding to sample outcomes with $r = 2, 3, 5, 7$, and $8$ ($\hat{p}_f \approx 2/50, 3/50, \ldots, 8/50$). Note that, because the distributions of $Z_{i,n}$ depend only on $\hat{p}_f \approx r/n$, we need not be concerned with the entire sample outcome when generating bootstrap distributions for this illustration. For $p_f = .5$, the distributions of $Z_{i,n}$ are similar for all values of $r$ (or $\hat{p}_f$), showing that PTBT works well in this case. For $p_f = .1$, the distributions of $Z_{i,n}$ are dissimilar among values of $r$, indicating that PTBT works poorly here. For $p_f = .2$, the agreement among the bootstrap distributions is good for $r \geq 5$ (or $\hat{p}_f > 5/50$) but not so good for smaller $r$, resulting in only moderately good behavior in this setting.

Note that the distribution of $Z_{i,n}$ at $r/n = 5/50 \approx \hat{p}_f$ has a highly discrete behavior. The reason for this...
can be seen by first noting that \( \delta_{\log t} \approx \sigma^* K \), where \( K \) is a constant. Combining this with (2) gives

\[
Z_{t_p} = \frac{\Phi^{-1}(p) - \Phi^{-1}(r/n)}{K}.
\]

(3)

When the approximation is good, the distribution of \( Z_{t_p} \) is approximately discrete, corresponding to the distribution of \( r \). As \( p_f \) moves away from \( p \), (2) and thus (3) are no longer good approximations, and the discrete-like behavior disappears. Figure 8 also shows that for values of \( r < n \times p_f \) the distribution of \( Z_{t_p} \) is strongly skewed to the left; for values of \( r > n \times p_f \) the distribution of \( Z_{t_p} \) is more symmetric.

Figure 9 shows bootstrap estimates of \( F^*(t) \) corresponding to sample outcomes \( r = 2, 5, 8 \) from Figure 6 [without loss of generality, the dark solid lines in Fig. 9 are taken to be the \( F(t) \) distribution from which the bootstrap sample is drawn], and correspond to the top, middle, and bottom rows of histograms in the \( p_f = .1 \) column of Figure 8. The mapping between Figure 9 and Figure 8 can be visualized by noting where the \( F^*(t) \) lines across the Proportion Failing = .1 line. The zero point on the distribution of \( \log(t^*_1) - \log(t_1) \) will correspond to the point where the \( F(t) \) line crosses the Proportion Failing = .1 line. For the \( r = 2 \) plot (where \( p_f = 2/50 < p = .1 \), the
\( \hat{F}^* (t) \) lines crossing where \( \log(t_{1}^*) - \log(t_{1}) > 0 \) tend to have very small slope (large \( \hat{\sigma}^{*} \)) values. This causes shrinking toward 0 of the \( Z_{i1}^* \) values and the corresponding left-skewed distribution for \( Z_{i1}^* \). For the \( r = 8 \) plot (where \( \hat{p}_f = 8/50 > p = .1 \)), the shrinking behavior is less pronounced and the result is the more symmetric distribution for \( Z_{i1}^* \).

Robinson (1983) used a parametric bootstrap procedure to find CI's for multiply time-censored progressive data. This procedure (similar to PTBT) is exact when data are complete or Type II censored. Because multiple time-censored data contain several censoring times, there is no discrete-like behavior in the MLE's like that seen with Type I censoring. For this reason the CP with multiple time censoring is close to the nominal over all of the different cases. For the Type I censored case with a single censoring point, however, our simulation results (details not shown here) showed that the coverage probabilities of Robinson's procedure tend to be less accurate than those of the PTBT procedure.

8. DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

Life tests usually result in Type I censored data. Because there are no known exact CI procedures for Type I censored data, this article provides a detailed comparison of procedures for constructing approximate CI's. These procedures range from the most commonly used large-sample normal approximation procedures to the more modern computationally intensive likelihood- and simulation-based procedures. Our results show that for moderate amounts of censoring and one-sided bounds (most commonly used in practical applications in the physical and engineering sciences as well as other areas of application) the simple normal-approximation (NORM and TNORM) procedures provide only crude approximations even when the expected number of failures is as large as 50 to 100.

Appropriate computationally intensive procedures provide important improvements. In particular, likelihood-based procedures, generally outperform the normal-approximation procedures. Calibrating the individual tails of a likelihood-based interval with simulation (i.e., the PBSRLLR procedure) provides further improvement in one-sided coverage probability accuracy, even for small \( E(r) \), for all but the exceptional case [i.e., inferences at times near to the censoring time or quantiles near the proportion censoring with \( E(r) \leq 10 \)]. The transformed bootstrap-\( t \) procedure provides a computationally simpler procedure, but one needs to be careful in the specification of the transformation to be used.

In addition to providing guidance for practical applications, our results suggest the following avenues for further research:

1. Our study leaves unanswered the question of what one should do when making inferences in the exceptional case when the expected number of failures is less than 10. We see no easy solution to this problem. Some possibilities include the following:
   - Extending the censoring time of the life test to be safely and sufficiently beyond the time point (or proportion failing) of interest. This requires prior knowledge of the failure-time distribution, which is not generally available.
   - Designing life-test experiments to result in Type II censored data. In this case, exact CI procedures are available, but experimenters generally have to deal with time constraints in life testing, and thus there may be resistance to such life-test plans. On the other hand, Type II censoring provides important control over the amount of information that a life-test experiment will provide.
   - Designing life-test experiments to result in multiple time censoring [in which the results of Robinson (1983) suggest that excellent large-sample approximations are available from computationally intensive procedures]. In this case, constraints on time or number of units available for testing may also lead to resistance to such life-test plans.
   - If none of the preceding is possible (e.g., for reasons given previously or because the experiment has already been completed), making use, if possible, of nonparametric methods (where conservative CI's or CB's may be available if there is a sufficient amount of data).

2. Our study has focused on the Weibull and lognormal distributions. We would expect very similar results for other log-location-scale distributions such as the loglogistic distribution and other censored-data situations that arise in applications, including regression analysis and the analysis of accelerated life-test data, more complicated censoring schemes like interval censoring and random censoring, simultaneous CI's and CB's, CI's to compare two different groups, and so on.

3. The LLRBART is second-order accurate for two-sided CI's using Type I censored data (Jensen 1993). Both PTBT and PBSRLLR procedures are better than LLRBART in one-sided cases. Simulation results also suggest that PBSRLLR is better than PTBT with smaller sample sizes. This finding suggests that higher-order asymptotics would show a difference between these different procedures. This could be explored.

4. As discussed in Section 4.1, our results are conditional on having a sample with at least two failures. When \( E(r) \) is small [e.g., \( E(r) < 10 \)], there can be a nonnegligible probability of having zero or one failure so that it is not possible to compute meaningful confidence intervals. A referee suggested that there might be some improvement in the performance of CI procedures by developing estimation procedures (including bootstrap) that explicitly condition on the fact that \( r \geq 2 \). It might be of interest to explore the use of such procedures.

ACKNOWLEDGMENTS

We thank Necip Doganaksoy, Luis A. Escobar, C. Joseph Lu, and Wayne Nelson for providing many useful sugges-
tions on an earlier version of this article. Comments provided by the referees, associate editor, and Editor Max Morris helped us to make important improvements in the article. Computing for the reported research was run on the DEC Alpha Farm of workstations at the Iowa State University Computation Center and equipment in the Department of Statistics, purchased with funds provided by an NSF SCREMS grant award DMS 9707740 to Iowa State University.

[Received December 1997. Revised May 1999.]

REFERENCES


