Comparisons of Approximate Confidence Interval Procedures for Type I Censored Data

Shuen-Lin Jeng
Dept. of Statistics
Ming-Chuan University
Kuei-Shan, Taoyuan 333
Taiwan R.O.C.

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

June 6, 1999

Abstract

This paper 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.

Keywords: 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 (CIs) are needed. For Type II (failure) censored data (or uncensored data) from location-scale distributions (or log-location-scale distributions), Lawless (1982, page 147) describes pivotal quantities that can be used to obtain exact CIs 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 small to moderate number of failures, especially for one-sided confidence bounds (CBs). We evaluate CI procedures in order to find those that have high accuracy for both one-sided CBs and two-sided CIs 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 ML estimators has a discrete component related to the random number of failures. Also, t-like quantities have distributions that depend on the p_f, the proportion failing at the censoring time. We show that, for these reasons, some bootstrap procedures behave poorly in constructing CIs for the p quantile when p is close to p_f and the expected number of failures is small.

1.2 Related Work

CIs 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 in Nelson 1982, pages 330-333) can improve the approximation to the normal distribution. For example, statistics transformed to have a range over whole real line may provide studentized (or t-like) statistics with distributions that are closer to normal than those with finite boundaries.

Piegorsch (1987) explored likelihood based intervals for two-parameter exponential samples with Type I censoring. For inference on the scale parameter, the coverage probabilities for two-sided CIs become adequate when the sample size reaches 25. Ostrouchov and Meeker (1988) showed that CIs based on inverting log likelihood ratio (LLR) tests provide a better approximation than TNORM CIs with interval censored data and Type I censoring for the Weibull and lognormal distributions. Vander Wiel and Meeker (1990) showed that for Type I censored Weibull data from accelerated life tests, LLR based CIs 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 CIs for reliability and life-data analysis applications. He notes that the LLRBART CIs 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 CIs. The parametric bootstrap procedure approximates the distribution of statistics by simulation or re-sampling. Hall (1987, 1992), Efron and Tibshirani (1993), Shao and Tu (1995) describe bootstrap theory and methods in detail.

Robinson (1983) applied a bootstrap procedure to location-scale distributions. The statistics used for constructing CIs are pivotal quantities in the case of complete or Type II censored data. He used the method to find CIs 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 CBs. 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 CIs that usually improves the performance of the percentile and BC procedures in complete samples.

The signed-root log-likelihood ratio (SRLLR) statistic has an approximate standard normal distribution in large samples (Barndorff-Nielsen and Cox 1994, page 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 CBs. To construct CIs that have approximately equal lower and the upper error probabilities, one can combine lower and upper one-sided CBs 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 paper is organized as follows. Section 2 describes the model and the estimation method. Section 3 provides details of the procedures for finding approximate CIs 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

$$F_Y(y) = 1 - \exp\left[-\exp\left(\frac{y-\mu}{\sigma}\right)\right],$$
$$-\infty < y < \infty, -\infty < \mu < \infty, \sigma > 0,$$

where μ and σ 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 $c_p = \log[-\log(1-p)]$ is the p quantile of the standardized ($\mu = 0$, and $\sigma = 1$) SEV distribution. The traditional Weibull scale and shape parameters are $\alpha = \exp(\mu)$ and $\beta = 1/\sigma$, respectively.

2.2 ML Estimation

We use $\widehat{\mu}$ and $\widehat{\sigma}$ to denote the ML estimators of the SEV parameters. Because of the invariance property of ML estimators, $\widehat{y}_p = \widehat{\mu} + c_p\widehat{\sigma}$ is the ML estimator of the p quantile of the SEV distribution. Also the ML estimators of the Weibull parameters are $\widehat{\alpha} = \exp(\widehat{\mu})$ and $\widehat{\beta} = 1/\widehat{\sigma}$. The ML estimator of the Weibull p quantile is $\widehat{t}_p = \exp(\widehat{y}_p)$. More generally the ML estimator of a function $\boldsymbol{g}(\mu,\sigma)$ is $\widehat{\boldsymbol{g}} = \boldsymbol{g}(\widehat{\mu},\widehat{\sigma})$. For any particular function of interest, it is possible to re-parameterize by defining a one-to-one transformation, $\boldsymbol{g}(\mu,\sigma) = (g_1(\mu,\sigma), g_2(\mu,\sigma)) = \boldsymbol{\theta}$, 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 above for (μ,σ). This provides a procedure for obtaining ML estimates and likelihood CIs for any scalar or vector function of (μ,σ). For more details see Lawless (1982, Chapter 4) or Meeker and Escobar (1998, Section 8.3.3).

Let $\boldsymbol{\theta} = (\theta_1, \theta_2)$ be the unknown parameter vector, where θ_1 is the parameter of interest and θ_2 is a nuisance parameter. Typically $\boldsymbol{\theta}$ could be (μ,σ) or (t_p,σ). We use $L(\boldsymbol{\theta})$ to denote the likelihood and t_c to denote the specified censoring time. Let $t_1, ..., t_n$ be n observations (e.g., failure or censoring times) from a life test. If the observations are independent, then the censored-data likelihood is

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

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

Table 1: Abbreviations for CI/CB procedures

NORM	Normal-approximation
TNORM	Transformed normal-approximation
LLR	Log likelihood ratio
LLRBART	Log likelihood ratio Bartlett corrected
PBT	Parametric bootstrap-t
PTBT	Parametric transformed bootstrap-t
PBSRLLR	Parametric bootstrap signed square root LLR
PBP	Parametric bootstrap percentile
PBBCA	Parametric bootstrap bias-corrected accelerated
PBBC	Parametric bootstrap bias-corrected

3 Confidence Interval/Bound Procedures

This section describes the different CI/CB procedures that are evaluated in this paper. Table 1 shows the abbreviation for each procedure. Let $C_{n;1-\alpha}$ denote an approximate CI for θ_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 kth order accurate if $\Pr(\theta_1 \in C_{n;1-\alpha}) = 1-\alpha+O(n^{-k/2})$. 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 $100(1-\alpha)\%$ confidence interval for each CI procedure used in the comparison. One-sided CBs are obtained by using the appropriate endpoint of a two-sided confidence interval, with a corresponding adjustment to the confidence level.

3.1 Normal-Approximation Procedures

Normal-approximation procedure (NORM). Suppose $\widehat{\boldsymbol{\theta}}$ is the ML estimator of the parameter vector $\boldsymbol{\theta}$. Under the usual regularity conditions, $\widehat{\boldsymbol{\theta}}$ is asymptotically normal and efficient (Serfling 1980, page 148). Let $I_{\boldsymbol{\theta}}$ denote the Fisher information matrix and let $n[\widehat{\text{se}}_{\widehat{\theta}_1}]^2$ be an estimator that converges to $I_{\boldsymbol{\theta}}^{(1,1)}$ in probability when n increases to ∞, where $I_{\boldsymbol{\theta}}^{(1,1)}$ is the $(1,1)$ term of the inverse of $I_{\boldsymbol{\theta}}$. Then the distribution of $(\widehat{\theta}_1-\theta_1)/\widehat{\text{se}}_{\widehat{\theta}_1}$ is approximately $N(0,1)$ in large samples. Thus a normal-approximation $100(1-\alpha)\%$ CI can be obtained from $\widehat{\theta}_1 \pm z_{(1-\alpha/2)}\widehat{\text{se}}_{\widehat{\theta}_1}$, where $z_{(1-\alpha/2)}$ is the $N(0,1)$ distribution $1-\alpha/2$ quantile. In this paper $n[\widehat{\text{se}}_{\widehat{\theta}_1}]^2$ is obtained from the inverse of the local estimate of $I_{\boldsymbol{\theta}}$ (e.g., Nelson 1982, page 377).

Transformed normal-approximation procedure (TNORM). When an ML estimator $\widehat{\theta}_1$ has its range on only part of the real line, a monotone function $g(\widehat{\theta}_1)$ with continuous derivatives and with range on the entire real line generally has a better normal-approximation (Nelson, 1982, page 331). Let $g'(\theta_1)$ denote the first derivative of $g(\theta_1)$ and let $n[\widehat{\text{se}}_{g(\widehat{\theta}_1)}]^2$ be an estimator that converges to $[g'(\theta_1)]^2 I_{\boldsymbol{\theta}}^{(1,1)}$ in probability. The TNORM procedure is based on the normal approximation $[g(\widehat{\theta}_1) - g(\theta_1)]/\widehat{\text{se}}_{g(\widehat{\theta}_1)} \dot{\sim} N(0,1)$. Then the TNORM CI procedure uses $g^{-1}[g(\widehat{\theta}_1) \pm z_{(1-\alpha/2)}\widehat{\text{se}}_{g(\widehat{\theta}_1)}]$, 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 paper $n[\widehat{\text{se}}_{g(\widehat{\theta}_1)}]^2$ is obtained, using the delta method, as $[g'(\widehat{\theta}_1)]^2 \widehat{I}_{\boldsymbol{\theta}}^{(1,1)}$, where $\widehat{I}_{\boldsymbol{\theta}}$ is the local estimate of $I_{\boldsymbol{\theta}}$.

3.2 Likelihood Ratio Procedures

Log LR procedure (LLR). The profile likelihood for θ_1 is defined as

$$R(\theta_1) = \max_{\theta_2} \left[\frac{L(\theta_1, \theta_2)}{L(\widehat{\boldsymbol{\theta}})} \right]. \tag{1}$$

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

Log LR Bartlett corrected procedure (LLRBART). Because the expectation of $W/\text{E}(W)$ is equal to the mean of the χ^2_1 distribution, the distribution of $W/\text{E}(W)$ will be better approximated by the χ^2_1 distribution (Bartlett 1937). In general one must substitute an estimate for $\text{E}(W)$ computed from one's data. For complicated problems (e.g., those involving censoring) it is necessary to estimate of $E(W)$ 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 $W(\theta_1)/\text{E}(W) - \chi^2_{(1-\alpha,1)} = 0$ 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 a statistic S is a function of random variables with a distribution that depends on the parameter $\boldsymbol{\theta}$. The parametric bootstrap version S^* of S is the same function but based on data ("bootstrap sample") simulated using $\widehat{\boldsymbol{\theta}}$ in place of the unknown $\boldsymbol{\theta}$. The distribution of S^* is easily obtained by simulation.

Parametric bootstrap-t procedure (PBT). (Efron 1982) Let $\widehat{\theta}_1$ be the ML estimator of θ_1 and let $\widehat{\theta}^*_1$ be the ML estimator from a bootstrap sample. Also let $z_{\widehat{\theta}^*_{1(\alpha)}}$ be the α quantile of the distribution of $Z_{\widehat{\theta}^*_1} = (\widehat{\theta}^*_1 - \widehat{\theta}_1)/\widehat{\text{se}}_{\widehat{\theta}^*_1}$, where $\widehat{\text{se}}_{\widehat{\theta}^*_1}$ is the bootstrap version of

$\widehat{\mathrm{se}}_{\widehat{\theta}_1}$. In this paper we choose $\widehat{\mathrm{se}}_{\widehat{\theta}_1}$ to be the same as in the NORM procedure. The PBT CI procedure uses $[\widehat{\theta}_1 - z_{\widehat{\theta}^*_{1(1-\alpha/2)}} \widehat{\mathrm{se}}_{\widehat{\theta}_1}, \quad \widehat{\theta}_1 - z_{\widehat{\theta}^*_{1(\alpha/2)}} \widehat{\mathrm{se}}_{\widehat{\theta}_1}]$.

Parametric transformed bootstrap-t procedure (PTBT). Let g be a smooth monotone function generally chosen such that $g(\widehat{\theta}_1)$ has range on whole real line. Let $\widehat{\theta}_1$ be the ML estimator of θ_1 and let $\widehat{\theta}^*_1$ be the bootstrap version ML estimator. Let $z_{g(\widehat{\theta}^*_1)_{(\alpha)}}$ be the α quantile of the distribution of $Z_{g(\widehat{\theta}^*_1)} = [g(\widehat{\theta}^*_1) - g(\widehat{\theta}_1)]/\widehat{\mathrm{se}}_{g(\widehat{\theta}^*_1)}$, where $\widehat{\mathrm{se}}_{g(\widehat{\theta}^*_1)}$ is the bootstrap version of $\widehat{\mathrm{se}}_{g(\widehat{\theta}_1)}$. In this paper we choose $\widehat{\mathrm{se}}_{g(\widehat{\theta}_1)}$ to be the same as in the TNORM procedure. When g is monotone increasing, the PTBT CI procedure for θ_1 uses $[g^{-1}\{g(\widehat{\theta}_1) - z_{g(\widehat{\theta}^*_1)_{(1-\alpha/2)}} \widehat{\mathrm{se}}_{g(\widehat{\theta}_1)}\}, \quad g^{-1}\{g(\widehat{\theta}_1) - z_{g(\widehat{\theta}^*_1)_{(\alpha/2)}} \widehat{\mathrm{se}}_{g(\widehat{\theta}_1)}\}]$. When g is monotone decreasing, the order of the endpoints is reversed.

Parametric bootstrap signed square root log LR procedure (PBSRLLR). Let $V(\theta_1) = \mathrm{sign}(\widehat{\theta}_1 - \theta_1)[-2\log R(\theta_1)]^{1/2}$ denote the signed square root of the log likelihood ratio statistic. In large samples, the distribution of $V(\theta_1)$ can be approximated by a normal distribution (Barndorff-Nielsen and Cox 1994, page 101). Approximating by simulation, however, captures the asymmetry of the distribution and hence provides a better approximation for finding CBs for θ_1. Suppose that $v_{\widehat{\theta}^*_{1(\alpha)}}$ is the α quantile of the bootstrap distribution of $V(\theta_1)$. Then, similar to the LLR procedure, the PBSRLLR CI procedure uses the roots of $V(\theta_1) - v_{\hat{\theta}^*{}_{1(\alpha/2)}} = 0$ and $V(\theta_1) - v_{\hat{\theta}^*{}_{1(1-\alpha/2)}} = 0$ as the lower and upper confidence bounds, respectively.

Parametric bootstrap percentile procedure (PBP), parametric bootstrap bias-corrected procedure (PBBC) and parametric bootstrap bias-corrected accelerated procedure (PBBCA). Efron (1981, 1982) described the PBP and PBBC procedures. Based on concerns expressed by Schenker and Patwardhan (1985), Efron (1987) suggested an improved percentile bootstrap procedure that corrected for both bias and non-constant scale and named it BCA (bias-corrected and accelerated) procedure. Efron and Tibshirani (1993, section 14.3) showed an easier way to obtain BCA CIs by using an easy-to-compute estimate of the acceleration constant.

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 a number of different reasons, including staggered entry of units into a study and multiple failure modes (see Section 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,

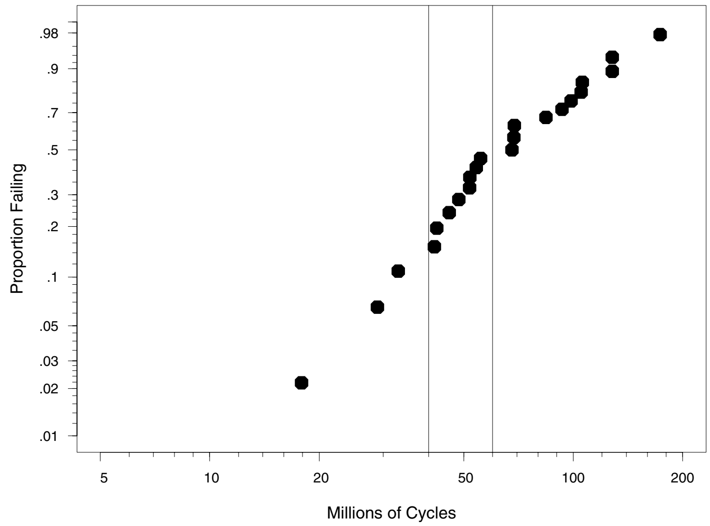

Figure 1: Probability plot for ball bearing fatigue data with vertical lines showing the points of artificial censoring.

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, page 228). Table 2 shows numerical values for two-sided approximate 90% confidence intervals (lower and upper one-sided approximate 95% confidence bounds) 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 the data censored at 40 million cycles. Also, the differences among the procedures are much more pronounced with the $t_c = 40$ million cycle data. The extremely wide intervals for $t_{.5}$ from the $t_c = 40$

Table 2: Comparison of confidence intervals for σ, $t_{.1}$, and $t_{.5}$, based on the the ball bearing data artificially censored at 40 and 60 million cycles.

Censoring time: 40 Million Cycles

Procedure	σ	$t_{.1}$	$t_{.5}$
NORM	[0.00, 0.89]	[17.97, 52.73]	[7.69, 148.92]
TNORM	[0.14, 1.28]	[21.62, 57.80]	[31.78, 192.95]
LLR	[0.17, 1.66]	[14.18, 82.04]	[47.90, 870.49]
LLRBART	[0.16, 1.78]	[17.35, 68.16]	[47.55, 1029.20]
PBT	[0.17, 2.09]	[17.22, 53.30]	[59.55, 367.14]
PTBT	[0.18, 2.02]	[20.48, 52.48]	[53.59, 1276.38]
PBSRLLR	[0.18, 2.07]	[19.41, 76.29]	[52.16, 2509.83]
PBP	[0.08, 1.00]	[21.83, 44.93]	[45.34, 236.71]
PBBC	[0.13, 1.24]	[22.80, 45.04]	[49.85, 690.15]
PBBCA	[0.14, 1.30]	[20.29, 44.21]	[47.55, 358.48]

Censoring time: 60 Million Cycles

Procedure	σ	$t_{.1}$	$t_{.5}$
NORM	[0.15, 0.50]	[21.46, 44.73]	[49.02, 72.95]
TNORM	[0.19, 0.56]	[23.29, 47.04]	[50.12, 74.20]
LLR	[0.20, 0.60]	[19.74, 43.26]	[50.76, 80.05]
LLRBART	[0.19, 0.61]	[19.18, 43.60]	[50.43, 81.12]
PBT	[0.21, 0.63]	[15.96, 43.65]	[53.67, 76.51]
PTBT	[0.20, 0.63]	[18.05, 43.30]	[52.98, 76.74]
PBSRLLR	[0.20, 0.63]	[18.72, 42.99]	[50.11, 81.60]
PBP	[0.17, 0.54]	[22.87, 46.46]	[50.73, 81.40]
PBBC	[0.18, 0.57]	[21.17, 44.64]	[50.40, 80.19]
PBBCA	[0.19, 0.62]	[18.31, 43.10]	[50.25, 79.84]

million cycles data are due to the large amount of extrapolation in time when estimating $t_{.5}$.

4 Simulation Experiment

This section describes the simulation experiment to compare the different CI/CB procedures.

4.1 Simulation Design

The simulation experiment was designed to study the effect of the following factors on coverage probability:

Table 3: Number of the cases where $r = 0$ or 1 in 2000 Monte Carlo simulations of the experiment. The expected numbers (rounded to the nearest integer) are shown inside parentheses.

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

- p_f: the expected proportion failing before the censoring time.
- $\text{E}(r) = np_f$: the expected number of failures before the censoring time.

We used 2000 Monte Carlo samples for each p_f and E(r) combination. The levels used were p_f = .01, .05, .1, .3, .5, .7, .9, 1 and E(r) = 3, 5, 7, 10, 15, 20, 30, 50, and 100. For each Monte Carlo sample, we obtained the ML estimates of the scale parameter and the quantiles y_p, p = .01, .05, .1, .3, .5, .632, and .9, where $\mu \cong y_{.632}$. The one-sided $100(1-\alpha)\%$ CBs were calculated for α =.025 and .05. Hence the two-sided CIs, 90% and 95%, can be obtained by combining the upper and lower CBs. Without loss of generality, we sampled from an SEV distribution with $\mu = 0$ and $\sigma = 1$.

The number of failures before the censoring time t_c is random. Therefore, it is possible to have as few as $r = 0$ or 1 failures in the simulation, especially when E(r) is small. 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 $r \leq 1$.

4.2 Coverage Probability Comparisons

Let $1 - \alpha$ be the nominal coverage probability (CP) of a CI procedure, and let $1 - \widehat{\alpha}$ denote the corresponding Monte Carlo estimate. The standard error of $\widehat{\alpha}$ is approximately $[\alpha(1-\alpha)/n_s]^{1/2}$, where n_s is the number of Monte Carlo simulation trials. For one-sided 95% CBs from 2000 simulations the standard error of the CP estimate is $[.05(1-.95)/2000]^{1/2} = .0049$. Thus the Monte Carlo error is approximately $\pm 1\%$. We say the procedure is adequate if the CP is within $\pm 2\%$ 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 (anti-conservative). We say that coverage probability is approximately symmetric when the CPs of the lower and upper CBs 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 failures. Those cases were excluded from coverage probability computation. With $\mathrm{E}(r) > 10$, there were no Monte Carlo simulations that had fewer than 2 failures.

5.1 One-sided CBs

Let UCB (LCB) denote an upper (lower) confidence bound. Figure 2 shows the coverage probability of the one-sided approximate 95% CBs for the parameter σ from 10 procedures for 5 cases of proportion failing. This figure shows that the TNORM procedure performs considerably 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 $\mathrm{E}(r) = 3$ case, dominating all of the other procedures evaluated here.

For estimating distribution quantiles, the situation is more complicated. Figure 3 gives CP versus $\mathrm{E}(r)$ for confidence interval procedures applied to the Weibull distribution quantile $t_{.1}$. As with σ, 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 $\mathrm{E}(r) \geq 15$. It also does well for $\mathrm{E}(r)$ as small as 3, except when estimating t_p 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 costs. 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 $\mathrm{E}(r) > 15$, 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. UCBs (LCBs) are conservative when $p < p_f$ ($p > p_f$) and are anti-conservative 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 CPs for bootstrap procedure for σ and several quantiles for $\mathrm{E}(r) = 15$, 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 CIs

As shown in Section 5.1, CPs 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.

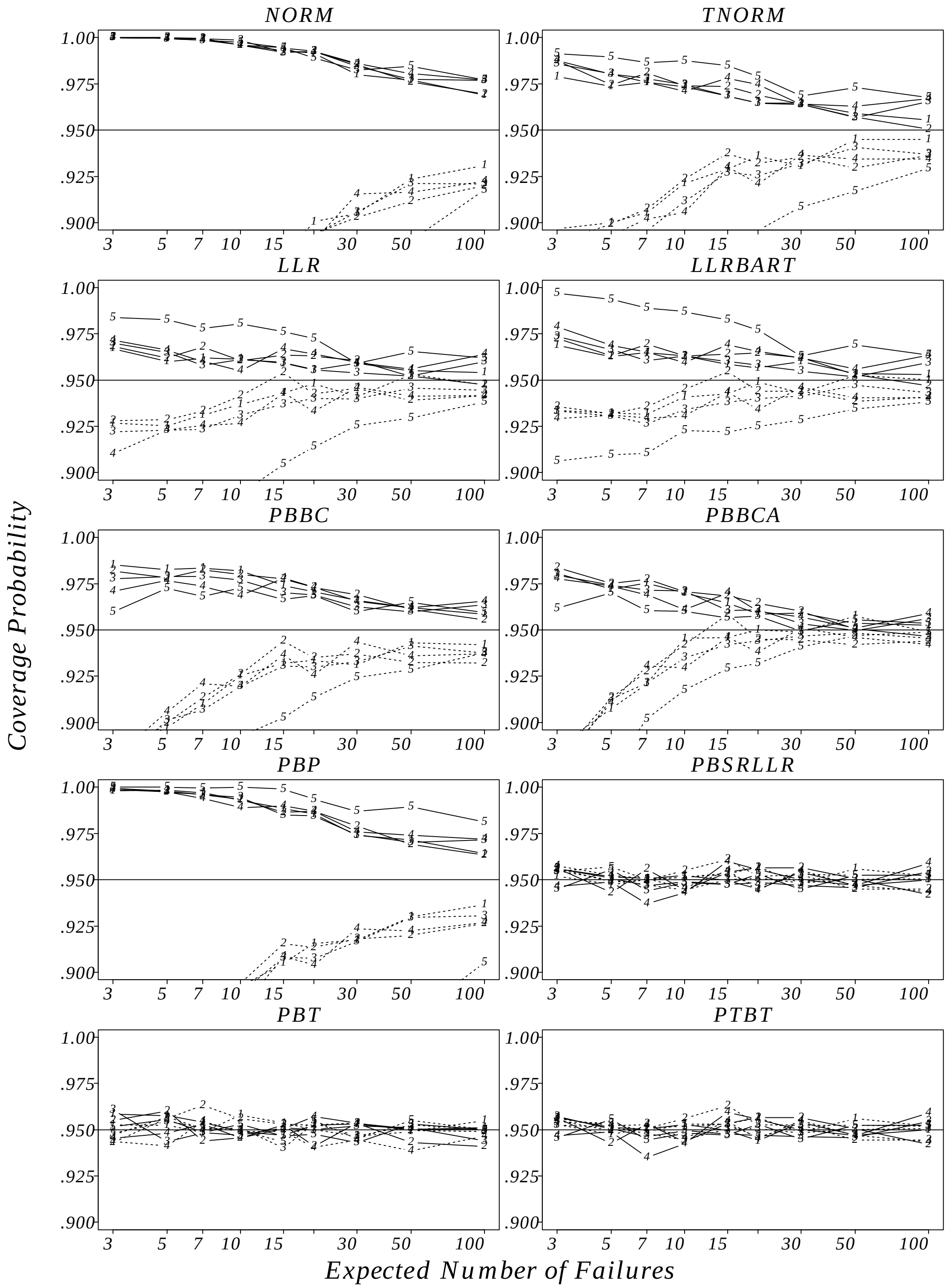

Figure 2: Coverage probability versus expected number of failures plot of one-sided approximate 95% CIs for parameter σ. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

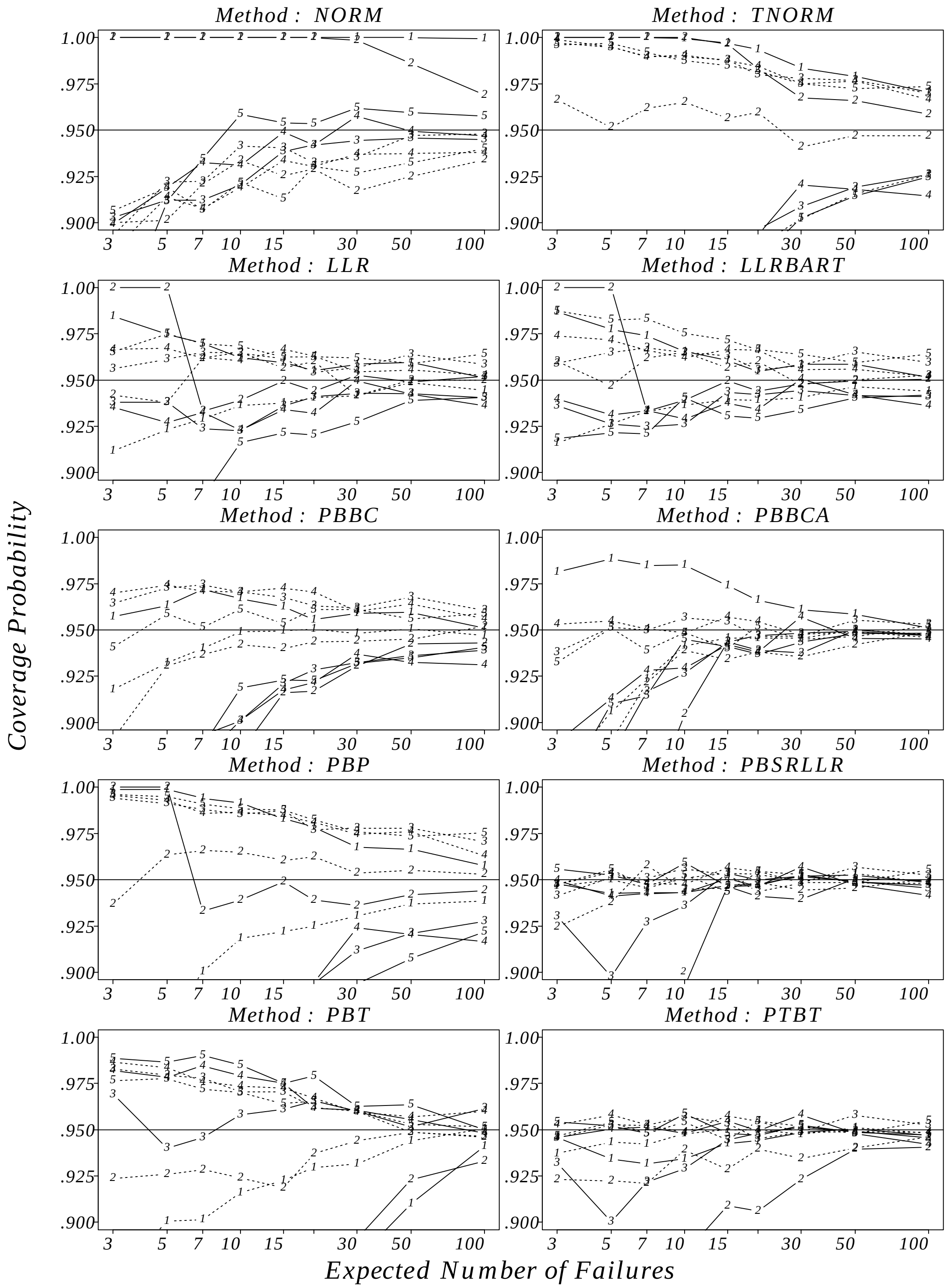

Figure 3: Coverage probability versus expected number of failures plot of one-sided approximate 95% CIs for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (.01, .1, .3, .5, 1). Dotted and solid lines correspond to upper and lower bounds, respectively.

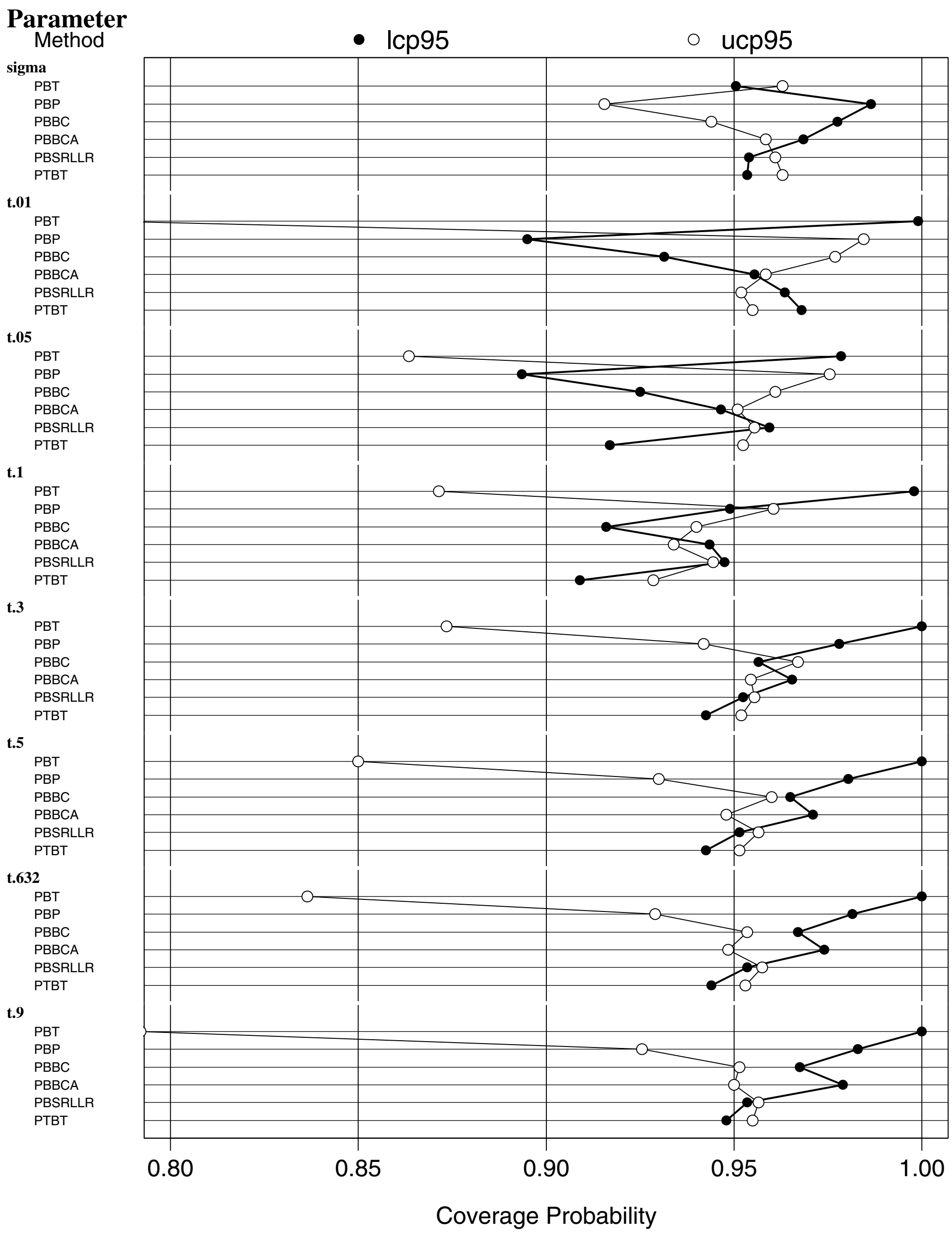

Figure 4: Coverage probability plot of approximate 95% one-sided CBs for bootstrap procedures in the case $E(r) = 15$ and $p_f = .1$.

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 σ 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) ≥ 15. Detailed results on confidence intervals for σ (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 CIs. With the same coverage probability, procedures that provide shorter intervals are better. Figures showing the mean interval length of the 2000 two-sided 90% CIs for parameters σ and $t_{.1}$ using 10 different procedures for 5 values of p_f can be found in Jeng (1998).

In comparing confidence interval widths (or, more precisely average width) it is preferable to compare intervals with nearly the same CP. Otherwise, procedures with conservative CPs tend to be wider than anti-conservative procedures (something that was easy to see in our results). When estimating σ, 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., Figures 10.5 and 10.6 in 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 CIs (NORM and TNORM), while 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 likelihood ratio procedures are consistent with results in Ostrouchov and Meeker (1988), Doganaksoy and Schmee (1993a), and

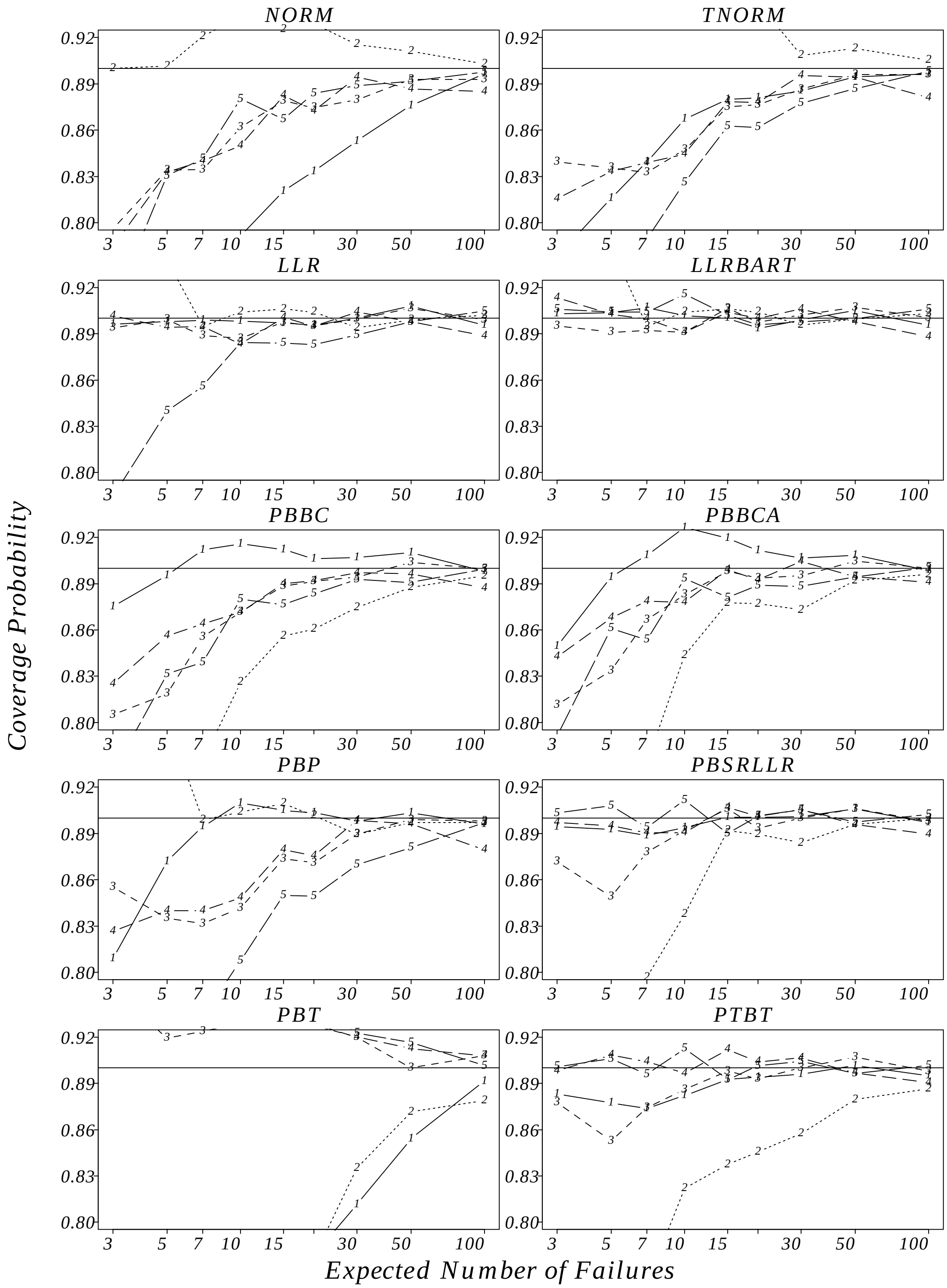

Figure 5: Coverage probability versus expected number of failures plot of two-sided 90% CIs for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (.01, .1, .3, .5, 1).

Doganaksoy (1995). This paper, however, focuses more on the asymmetry of coverage probability for one-sided CIs as well as cases with heavy censoring and a small expected number of failures.

Some bootstrap procedures provide better coverage probability accuracy. However, using the bootstrap-t without a proper transformation 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 procedure 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 number of failures, proportion failing and parameters of interest except for the case that 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 CBs 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 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 likelihood ratio based procedures can be recommended for finding one-sided CBs (two-sided CIs). For smaller E(r), the PTBT and PBSRLLR procedures can be recommended except for the case when the quantity of interest is t_p where p is close to proportion failing. Then PBSRLLR is better than PTBT down to $E(r)$=15. When $p_f > .5$, the PBSRLLR provides accurate CP even down to $E(r) = 10$. With modern computing capabilities, the PBSRLLR procedure is feasible and, when appropriate software becomes available, should be considered the best practice.

7 Special Effects of Type I Censoring

This section describes some of the special properties of ML estimators and related CI procedures that arise with Type I censored data. We then show how these properties relate to the exceptional cases where the bootstrap procedures do not perform well.

Doganaksoy and Schmee (1993b) explain that when the parameter of interest is t_p and p is close to the proportion failing p_f, then the ML estimates of t_p and σ are approximately uncorrelated. They go on to say that the TNORM procedure benefits from this property in that the error probabilities in the tails are more symmetric (but recall that both their results and our results show that an expected number of failures on the order of 50 to 100 is needed in order to have an adequate approximation to the nominal coverage probability). As we have shown, CIs calibrated with bootstrap/simulation provide an extremely good approximation with a large to moderately large expected number of failures. Interestingly, with a

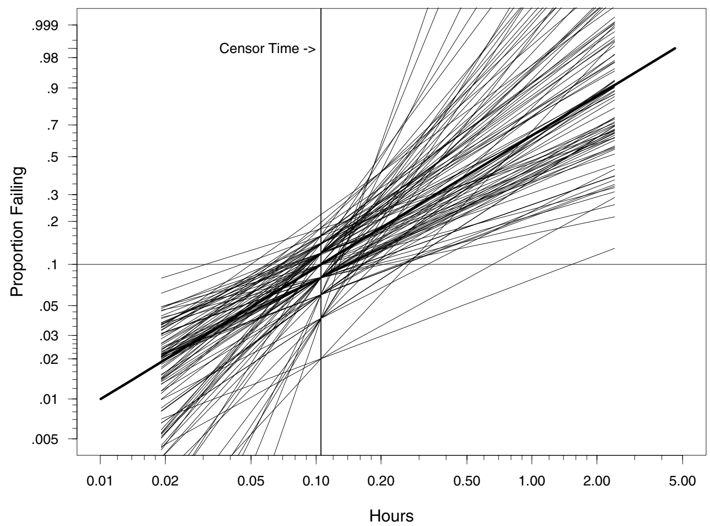

Figure 6: Weibull distribution life test simulation. The results show 100 ML estimates $\widehat{F}(t)$ based on data simulated from a Weibull distribution with parameters $\mu = 0$, $\sigma = 1$, sample size $n = 50$, with Type I censoring at a time corresponding to $p_f = .1$ so $\mathrm{E}(r) = 5$.

small expected number of failures, the CI procedures calibrated with bootstrap/simulation provide a good approximation *except* when t_p is to be estimated and p is close to the proportion failing p_f. This section describes and illustrate the reasons for this exceptional case for the PTBT procedure for estimating $t_{.1}$. The explanation would be similar for the other bootstrap procedures and quantiles.

As noted earlier, for complete data and Type II censoring, the distribution of the t-like statistics for quantiles like

$$Z_{\widehat{t_p}} = \frac{\log(\widehat{t}_p) - \log(t_p)}{\widehat{\mathrm{se}}_{\log(\widehat{t}_p)}}$$

and $Z_{\widehat{t_p^*}}$, the corresponding bootstrap version, are pivotal. In this case, the PTBT procedure is exact. For Type I censoring, however, the distribution of $Z_{\widehat{t_p}}$ depends on the sample size n and unknown p_f (correspondingly, the distribution of $Z_{\widehat{t_p^*}}$ depends directly on $\widehat{p}_f$ and thus indirectly on p_f). An outline of the proof of this result is given in Exercise 9.5 of Meeker and Escobar (1998). The distributions of $Z_{\widehat{t_p}}$ and $Z_{\widehat{t_p^*}}$ are similar. Because it is used for the PTBT method, the following discussion focuses primarily on $Z_{\widehat{t_p^*}}$.

As will be illustrated below, the distribution of $Z_{\widehat{t_p^*}}$ does not depend strongly on $\widehat{p}_f$, except when p is close to p_f. In the exceptional case, when p is close to p_f, the dependency of the distribution of $Z_{\widehat{t_p^*}}$ on $\widehat{p}_f$ is strong, causing poor performance. The rest of this section describes some of the details needed to understand this behavior.

Figure 6 shows the results of a simulation of 100 life tests with $\mu = 0$, $\sigma = 1$, and censoring time t_c chosen such that $p_f = .1$. Note the strong clustering of the ML estimate

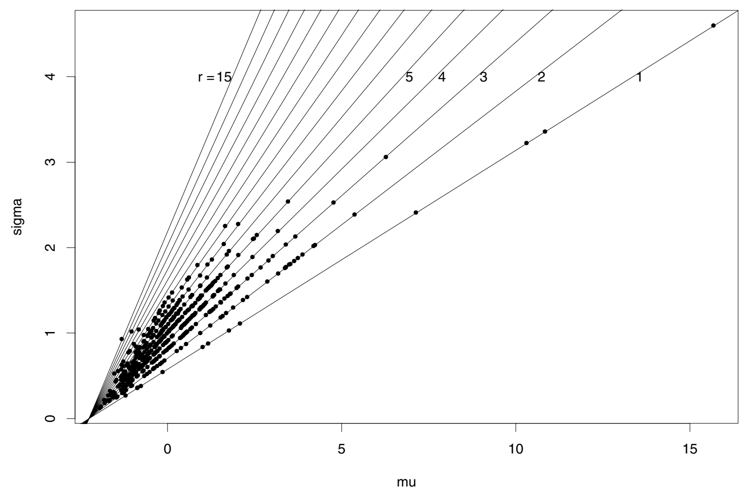

Figure 7: Plotted points are values of $(\widehat{\mu}, \widehat{\sigma})$ from 500 ML estimates from samples of size 50 with $p_f = .1$ [so $\mathrm{E}(r) = 5$] from the $\mu = 0$, $\sigma = 1$ Weibull distribution. The lines are $\mu + \Phi^{-1}(r/50)\sigma = \log(t_{.1})$ for $r = 1, 2, \ldots, 15$.

lines as they cross the vertical line at the censoring time $t_c = -\log(1 - .1)$. This clustering is due to the fact that $\widehat{p}_f = \widehat{F}(t_c) \approx r/n$. Then, as seen in Figure 6, and as can be shown analytically, the approximation is excellent for small r. This approximation implies that

$$\widehat{\mu} \approx \log(t_c) - \Phi^{-1}(r/n)\widehat{\sigma}. \tag{2}$$

Thus for a given number of failures $r \geq 1$, the estimates $\widehat{\mu}$ and $\widehat{\sigma}$ are almost exactly linearly related. This is shown in Figure 7. Figure 7 also shows the strong positive correlation between $\widehat{\mu}$ and $\widehat{\sigma}$, typical of heavy right censoring.

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

Note that the distribution of $Z_{\widehat{t_p^*}}$ at $r/n = 5/50 \approx \widehat{p}_f \approx p_f = .1$ has a highly discrete behavior. The reason for this can be seen by first noting that $\widehat{\mathrm{se}}_{\log(\widehat{t_p^*})} \approx \widehat{\sigma}^* K$ where K is a

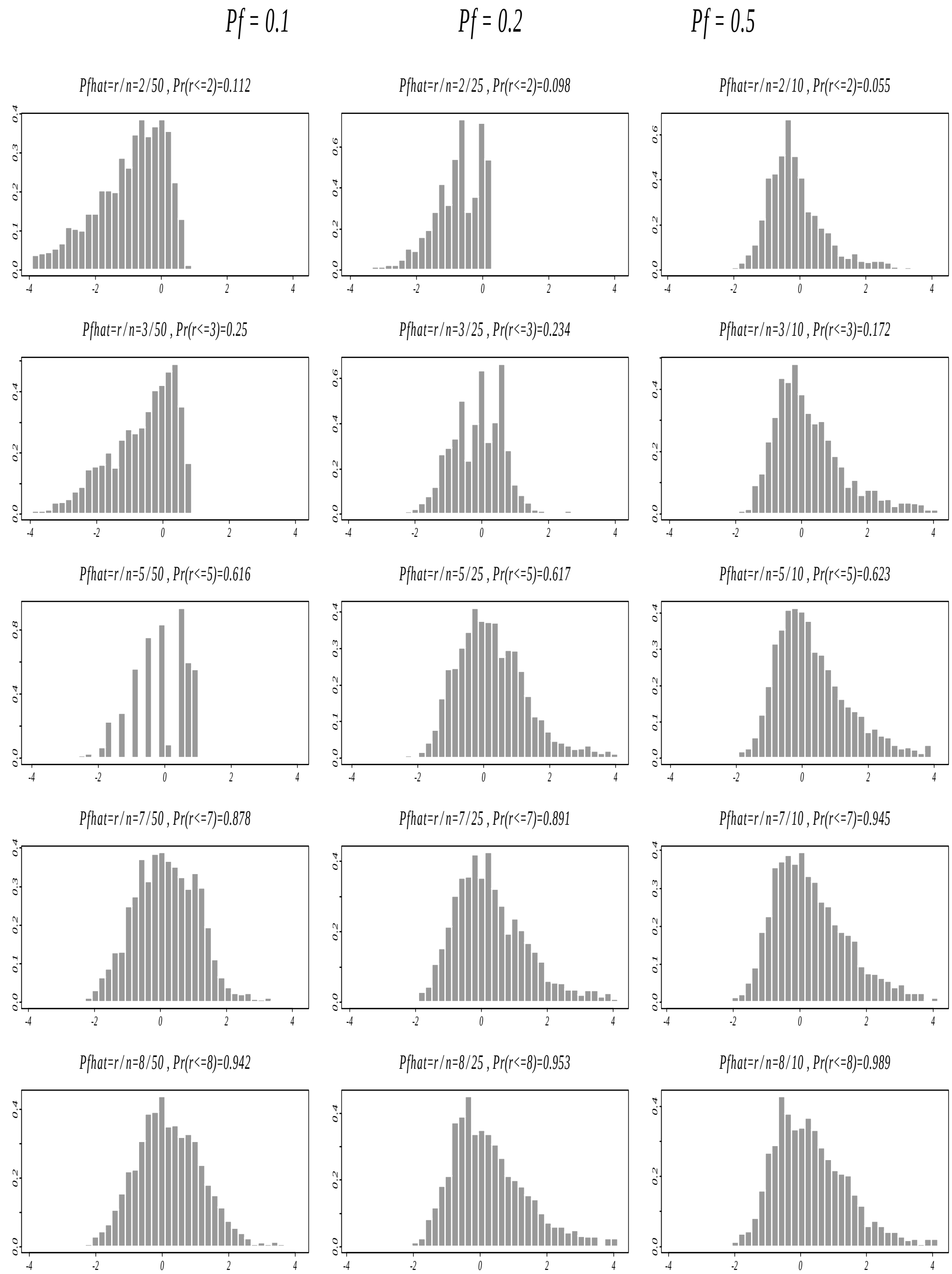

Figure 8: Distributions of the t-like statistic $Z_{\widehat{t}^*_{.1}} = [\log(\widehat{t}^*_{.1}) - \log(\widehat{t}_{.1})]/\widehat{\text{se}}_{\log(\widehat{t}^*_{.1})}$ for the PTBT procedure under Type I censoring with the same censoring time for true $p_f = .1, .2, .5$ and $E(r) = 5$. Each histogram is obtained by using 2000 simulations and is presented for those failure numbers > 1.

constant. Combining this with (2) gives

$$Z_{\widehat{t_p^*}} \approx \frac{\Phi^{-1}(p) - \Phi^{-1}(r/n)}{K}. \tag{3}$$

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

Reasons for the behavior in Figure 8 can be seen in Figure 9. Figure 9 shows bootstrap estimates of $\widehat{F}^*(t)$ corresponding to sample outcomes $r = 2, 5, 8$ from Figure 6 [without loss of generality the dark solid lines in Figure 9 are taken to be the $\widehat{F}(t)$ 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 $\widehat{F}^*(t)$ lines cross the Proportion Failing = .1 line. The zero point on the distribution of $\log(\widehat{t}^*_{.1}) - \log(\widehat{t}_{.1})$ will correspond to the point where the $\widehat{F}(t)$ line crosses the Proportion Failing = .1 line. For the $r = 2$ plot (where $\widehat{p}_f = 2/50 < p = .1$), the $\widehat{F}^*(t)$ lines crossing where $\log(\widehat{t}^*_{.1}) - \log(\widehat{t}_{.1}) > 0$ tend to have very small slope (large $\widehat{\sigma}^*$) values. This causes shrinking toward 0 of the $Z_{\widehat{t^*_{.1}}}$ values and the corresponding left-skewed distribution for $Z_{\widehat{t^*_{.1}}}$. For the $r = 8$ plot (where $\widehat{p}_f = 8/50 > p = .1$), the shrinking behavior is less pronounced and the result is the more symmetric distribution for $Z_{\widehat{t^*_{.1}}}$.

Robinson (1983) used a parametric bootstrap procedure to find CIs for multiply time-censored progressive data. This procedure (similar to PTBT) is exact when data are complete or Type II censored. Since multiple time-censored data contain several censoring times, there is no discrete-like behavior in the MLEs 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 paper provides a detailed comparison of procedures for constructing approximate CIs. 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 out-perform the normal-approximation

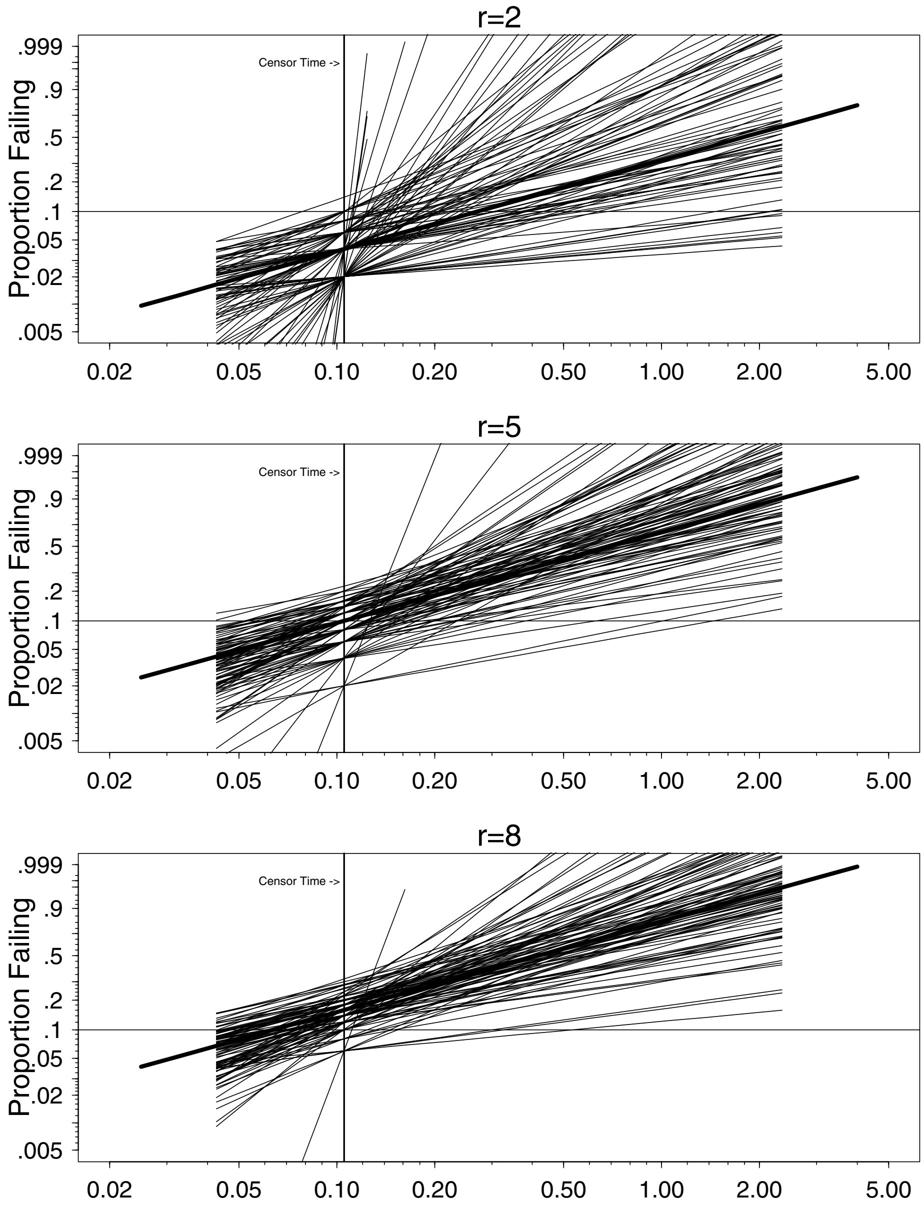

Figure 9: Plots of 100 bootstrap cdf estimates with Type I censoring at the same censoring time for sample estimates $\widehat{p}_f \approx 2/50$, $5/50$, and $8/50$.

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

    - 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.

    - Design 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.

    - Design life test experiments to result in multiple time-censoring (where 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 above is possible (e.g., for reasons given above or because the experiment has already been completed) it might be possible to make use of nonparametric methods (where conservative CIs or CBs 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 distribution 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 CIs and CBs, CIs to compare two different groups, and so on.

3. The LLRBART is second-order accurate for two-sided CIs 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 $\mathrm{E}(r)$ is small (e.g. $\mathrm{E}(r) < 10$), there can be a non-negligible 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 confidence interval 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 would like to thank Necip Doganaksoy, Luis A. Escobar, C. Joseph Lu and Wayne Nelson for providing many useful suggestions on an earlier version of this paper. Comments provided by the referees, Associate Editor, and Editor Max Morris helped us to make important improvements in the paper. Computing for the research reported in this paper was run on the DEC Alpha Farm of workstations at the Iowa State University Computation Center and equipment in Department of Statistics, purchased with funds provided by an NSF SCREMS grant award DMS 9707740 to Iowa State University.

References

Barndorff-Nielsen, O. E. (1986), "Inference on full or partial parameters based on the standardized signed log likelihood ratio," *Biometrika*, **73**, 307-322.

Barndorff-Nielsen, O. E. (1991), "Modified signed log likelihood ratio," *Biometrika*, **78**, 557-563.

Barndorff-Nielsen, O. E., and Cox, D. R. (1994), *Inference and Asymptotics*, London: Chapman & Hall

Bartlett, M. S. (1937), "Properties of sufficiency and statistical tests," *Proceedings of Royal Society of London* A, **160**, 268-282.

Doganaksoy, N. (1995), "Likelihood ratio confidence intervals in life-data analysis," Chapter 20 of *Recent Advances in Life-Testing and Reliability*, N. Balakrishnan (Editor), Boca Raton: CRC Press.

Doganaksoy, N., and Schmee, J. (1993a), "Comparisons of approximate confidence intervals for distributions used in life-data analysis," *Technometrics*, **35**, 175-184.

Doganaksoy, N., and Schmee, J. (1993b), "Orthogonal parameters with censored data," *Communications in Statistics, Part A Theory and Methods*, **22**, 669-685.

Efron, B. (1981), "Nonparametric standard errors and confidence intervals (with discussion)," *The Canadian Journal of Statistics*, **9**, 139-172.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia: Society for Industrial and Applied Mathematical.

Efron, B. (1987), "Better bootstrap confidence intervals (with discussion)," *Journal of the American Statistical Association*, **82**, 171-200.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap,* New York: Chapman & Hall.

Hall, P. (1987), "On the bootstrap and likelihood-based confidence regions," *Biometrika*, **74**, 481-493.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Jeng, S. L. (1998), "Improved Approximate Confidence Intervals for Censored Data," Ph.D. dissertation, Iowa State University.

Jensen, J. L. (1993), "A historical sketch and some new results on the improved log likelihood ratio statistic," *Scandinavian Journal of Statistics. Theory and Applications*, **20**, 1-15.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data,* New York: John Wiley & Sons.

Lawless, J. F. (1983), "Statistical methods in reliability," *Technometrics* **25**, 305-316.

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

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

Ostrouchov, G., and Meeker, W. Q., Jr. (1988), "Accuracy of approximate confidence bounds computed from interval censored Weibull and lognormal data," *Journal of Statistical Computation and Simulation*, **29**, 43-76.

Piegorsch, W. W. (1987), "Performance of likelihood-based interval estimates for two-parameter exponential samples subject to Type I censoring," *Technometrics*, **29**, 41-49.

Robinson, J. A. (1983), "Bootstrap confidence intervals in location-scale models with progressive censoring," *Technometrics*, **25**, 179-187.

Schenker, N., and Patwardhan, G. (1985), "Qualms about bootstrap confidence intervals," *Journal of the American Statistical Association*, **80**, 360-361.

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons.

Shao, J., and Tu, D. (1995), *The Jackknife and Bootstrap*, New York: Springer-Verlag.

Vander Wiel, S. A., and Meeker, W. Q. (1990), "Accuracy of approximate confidence bounds using censored Weibull regression data from accelerated life tests," *IEEE Transactions on Reliability*, **R-39**, 346-351.