Asymptotic Properties of Bootstrap Likelihood Ratio Statistics for Time Censored Data

Shuen-Lin Jeng
Dept. of Statistics
Tunghai University
Taichung, 407
Taiwan, R. O. C.

S.N. Lahiri[†]
Dept. of Statistics
Iowa State University
Ames, IA 50011
U. S. A.

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

February 17, 2003

Abstract

Much research has been done on the asymptotic distributions of likelihood ratio statistics for complete data. In this paper we consider the situation in which the data are censored and the distribution of the likelihood ratio statistic is a mixture of continuous and discrete distributions. We show that the distribution of signed square root likelihood ratio statistic can be approximated by its bootstrap distribution up to second order accuracy. Similar results are shown to hold for likelihood ratio statistics with or without a Bartlett correction. The main tool used is a continuous Edgeworth expansion for the likelihood-based statistics, which may be of some independent interest. Further, we use a simulation study to investigate the adequacy of the approximation provided by the theoretical result by comparing the finite-sample coverage probability of several competing confidence interval procedures based on the two parameters Weibull model. Our simulation results show that, in finite samples, the methods based on the bootstrap signed square root likelihood ratio statistic outperform the bootstrap-t and BC_a methods in constructing one-sided confidence bounds when the data are Type I censored.

Keywords: Bartlett correction, confidence interval, life data, likelihood ratio, one-sided confidence bound, signed square root likelihood ratio, parametric bootstrap, Type I censoring.
AMS(200) Subject Classification: Primary 62G09, Secondary 62G20.
[†] Research partially supported by NSF grant number DMS 0072571.

1 Introduction

The asymptotic distributions of likelihood ratio statistics have been studied for decades. Most previous work has focused on the situations where the data are given by a random sample of *complete* observations from a continuous distribution. For censored data, approximations to the distributions of likelihood ratio statistics are less well studied. A major technical problem in generalizing the results for the complete data case to time-censored data is that under censoring, the log-likelihood function and its derivatives become a mixture of partly discrete and partly continuous random variables, making the derivation of the relevant Edgeworth expansions difficult. In a series of important papers, Jensen (1987, 1989, 1993) developed Edgeworth expansions of the log-likelihood ratio (LLR) statistic when the underlying distribution is partly discrete. Because of the more complicated nature of the resulting expansions, however, the accuracy of certain likelihood-based procedures, that are second and higher order accurate in the complete data case, remains unexplored in the censored data case. In this paper, we investigate higher-order properties of some of these inference procedures under censoring and also investigate accuracy of bootstrap approximations to many common likelihood-based procedures under censoring.

The main results of the paper give continuous Edgeworth expansions of a general order for the multivariate maximum likelihood estimators (MLEs) under time censoring (also known as type I censoring). Validity of continuous third order expansions for the likelihood ratio statistic, its Bartlett corrected version, and the signed square root likelihood ratio statistic are also established. Using the continuous Edgeworth expansion results, we study the accuracy of approximations generated by a parametric bootstrap method. It is shown that if the MLE is Studentized using the Cholesky decomposition of a consistent estimator of the Fisher information matrix, the bootstrap approximation to the distribution of the multivariate Studentized MLE is second order accurate. Thus, the superiority of the bootstrap continues to hold for censored data, although under censoring the likelihood function and its partial derivatives involve discrete variables arising from the random number of failures of time censoring. One-term Edgeworth correction by the bootstrap is also established for the likelihood ratio statistic and its variants, confirming its superiority over the classical χ^2- and normal approximations. We also carry out an extensive simulation study to investigate the finite sample properties of the parametric bootstrap method under censoring. We consider a number of different likelihood-based approaches for constructing confidence intervals and study the accuracy of coverage probabilities for one- and two-sided confidence intervals as a function of sample size and the expected number of failures. The bootstrap-t and BC_a methods are known to be second order accurate when the data are complete (cf. Hall (1992)). Our simulation results show that the methods based on bootstrap signed square root likelihood ratio statistics outperform the bootstrap-t and BC_a methods in constructing one-sided confidence bounds when the data are time (or Type I) censored. For the two sided confidence intervals, the bootstrap signed square root likelihood ratio statistics has the best performance.

We conclude this section with a brief literature review. For independent and identically distributed (i.i.d.) complete data, Box (1949) derives an infinite series expansion for the distribution of the LLR statistic W_n (say) in terms of the χ^2 distribution and with terms decreasing in powers of $1/n$. Lawley (1956) derives the Bartlett correction term for W_n.

Doganaksoy and Schmee (1993) compares several confidence interval procedures using the W_n and its Studedized modifications. Chandra and Ghosh (1979) derives a valid Edgeworth expansion for W_n to order $o(1/n)$. For the signed root LLR statistic R_n (say), expansions for different versions of R_n have been derived by Lawley (1956), McCullagh (1984), Efron (1985) and Nishii and Yanagimoto (1993). In two important papers, Barndorff-Nelson (1986, 1991) show that a particular modification of R_n is asymptotically normal up to an error of the order $O(n^{-3/2})$ conditionally on an appropriate ancillary, and hence also unconditionally.

For censored data, the usual arguments for finding a formal Edgeworth expansion are no longer valid. The order of accuracy in the results mentioned above could be different. Jensen (1987, 1989, 1993) establish Edgeworth expansions for smooth functions of the mean when the underling distribution is partly discrete. These expansions are used to prove the validity of expansions for W_n. Babu (1991) establishes Edgeworth expansions for statistics that are functions of lattice and non-lattice variables for the case that the lattice variable is only one dimensional.

A large number of bootstrap methods have been suggested for testing or finding confidence intervals (Hall 1992, Efron and Tibshirani 1993, Shao and Tu 1995). The theoretical arguments for the accuracy of these methods are mostly derived under the assumption of complete data. For time-censored data, observation stops at a predetermined point in time. In this case, some bootstrap methods can be much less accurate, especially for one-sided confidence intervals and small expected number of failures (see Jeng and Meeker, 2000). Datta (1992) establishes a continuous version of classical Edgeworth expansions for both non-lattice and lattice distributions and uses this to unify both non-parametric and parametric bootstrap methods of a Studentized statistic up to order $O(n^{-1/2})$. Datta (1992) also gives an example that bootstrap-t method is first order accurate for the Type I censored data with the exponential distribution.

The rest of the paper is organized as follows. In Section 2, we briefly describe the theoretical framework and the bootstrap method. In Section 3, we derive continuous Edgeworth expansions for several likelihood-based statistics and in Section 4, we use these expansions to study higher order properties of the bootstrap approximations. In Section 5, we present a simulation study. In Section 6, we summarize the results and give some possible areas for future research. Proofs of the main results are given in Section 7.

2 Theoretical Framework

2.1 Likelihood-based Statistics

Let $X_1, X_2, \ldots$ be a sequence of $\mathbb{R}^d$ valued independent and identically distributed (i.i.d.) random vectors with common distribution P_θ, where θ belongs to an open subset Θ of $\mathbb{R}^k$. Suppose that P_θ is absolutely continuous w.r.t some σ-finite measure μ with density $f(x;\theta)$. Denote the cumulative distribution function (cdf) of P_θ by $F(x;\theta)$. With single Type I censoring at censor time t_c, the log-likelihood of a single observation is given by

$$l(X_i;\theta) = \log\{f(X_i;\theta)^{\delta_i}[1 - F(X_i;\theta)]^{1-\delta_i}\}, \tag{2.1}$$

where $\delta_i = 1$, if $X_i \le t_c$ (a failure) and $\delta_i = 0$, if $X_i > t_c$ (a censored observation), $i = 1, \ldots$. When there are n observations, define $\bar{l}_n$ as

$$\bar{l}_n(\theta) = \frac{1}{n}\sum_{1}^{n} l(x_i;\theta), \tag{2.2}$$

where x_i is the data for the observation i. Let $\widehat{\theta}_n = (\widehat{\theta}_{1n}, \ldots, \widehat{\theta}_{kn})$ be the maximum likelihood estimates of the parameter θ. Then, $\widehat{\theta}_n$ satisfies the k equations

$$\frac{\partial}{\partial\theta_i}\bar{l}_n(\theta) = 0, \quad i = 1, \ldots, k. \tag{2.3}$$

Next, let $\theta = (\theta^{(1)}, \theta^{(2)}) = (\theta_1, \ldots, \theta_{k_1}, \theta_{k_1+1}, \ldots, \theta_k)$, be a partition of the parameter vector θ where $\theta^{(2)}$ is the parameter of primary interest and $\theta^{(1)}$ is a vector of nuisance parameters, and let $\theta_0 = (\theta_0^{(1)}, \theta_0^{(2)}) = (\theta_{10}, \ldots, \theta_{k_1 0}, \theta_{(k_1+1)0}, \ldots, \theta_{k0})$, be the true parameter vector. Let $\tilde{\theta}_n = (\tilde{\theta}_{1n}, \ldots, \tilde{\theta}_{k_1 n}, \theta_{(k_1+1)0}, \ldots, \theta_{k0}) = (\tilde{\theta}_n^{(1)}, \theta_0^{(2)})$ be the maximum likelihood estimate of θ under the restricted model $\theta^{(2)} = \theta_0^{(2)}$. Then the log likelihood ratio statistic is

$$W_n \equiv W_n(\theta_0; k_1) = 2n[\bar{l}_n(\widehat{\theta}_n) - \bar{l}_n(\tilde{\theta}_n)], \tag{2.4}$$

and under standard regularity conditions (e.g., Lehmann 1986), the distribution of W_n is asymptotically $\chi^2_{(k-k_1)}$, where χ^2_f denotes a chi-square distribution with degree of freedom f.

The distribution of a likelihood ratio statistic with a Bartlett adjustment can be more closely approximated by the chi-square approximation than the distribution of a likelihood ratio statistic without a Bartlett adjustment. Consider the modified statistic

$$W_{1n} \equiv W_{1n}(\theta_0; k_1) = (k - k_1)\frac{W_n}{\mathrm{E}_{\theta_0}(W_n)} \tag{2.5}$$

and an expansion of $\mathrm{E}_{\theta_0}(W_n)$

$$\mathrm{E}_{\theta_0}(W_n) = (k - k_1)\left[1 + \frac{B(\theta_0)}{n}\right] + O\left(\frac{1}{n^2}\right). \tag{2.6}$$

Then, operationally, a Bartlett adjusted statistic WB_n can be obtained by

$$WB_n \equiv WB_n(\theta_0; k_1) = \frac{W_n}{1 + B(\tilde{\theta}^{(1)}, \theta_0^{(2)})/n}, \tag{2.7}$$

where $(\tilde{\theta}^{(1)}, \theta_0^{(2)})$ is the maximum likelihood estimate for the model parameter $\theta^{(1)}$ with the restriction $\theta^{(2)} = \theta_0^{(2)}$.

The signed square root log likelihood ratio (SRLLR) statistic for testing a scalar parameter $\theta_0^{(2)} = \theta_{k0}$ (or a scalar function of the parameter so that $k_1 = k - 1$) is

$$R_n \equiv R_n(\theta_0; k_1) = \mathrm{sign}(\widehat{\theta}_{kn} - \theta_{k0})\sqrt{W_n}, \tag{2.8}$$

and the distribution of R_n is asymptotically standard normal.

2.2 Bootstrap

Let $\bar{\theta}_n$ be an estimator of the parameter θ. For example, we may take $\bar{\theta}_n = \hat{\theta}_n$, the MLE of θ. Then, given the data $X_1, \ldots, X_n$, draw a random sample $X_1^*, \ldots, X_n^*$ of size n from the "estimated" density $f(x; \bar{\theta}_n)$. Then, for a random variable $T_n \equiv t_n(X_1, \ldots, X_n; \theta)$, define the *parametric* bootstrap version of T_n as

$$T_n^* = t_n(X_1^*, \ldots, X_n^*; \bar{\theta}_n). \tag{2.9}$$

In absence of a parametric model, the bootstrap samples may be drawn with replacement from the observations $\{X_1, \ldots, X_n\}$. The corresponding method is known as the"ordinary" bootstrap or the nonparametric bootstrap. Several authors have investigated properties of the nonparametric bootstrap for censored data. See Lo and Singh (1986), Horvath and Yandell (1987), Babu (1991), Lai and Wang (1993), Gross and Lai (1996), and the references therein. In this paper, we will consider the parametric bootstrap, rather than the nonparametric bootstrap.

3 Continuous Edgeworth Expansions

In this section, we derive Edgeworth expansions for the likelihood-based statistics of Section 2.1 by allowing the underlying parameter value to depend on the sample size. This approach has been introduced in the bootstrap literature by Datta (1992) and seems to be the most natural one for studying higher order properties of the parametric bootstrap method of Section 2.2. Let $\theta_0 \in \Theta$ and let $\{\theta_n\}_{n \geq 1} \subset \Theta$ be a sequence of parameter values satisfying

$$\theta_n \to \theta_0 \quad \text{as} \quad n \to \infty. \tag{3.1}$$

Also, write E_n and P_n to denote the expectation and the probability under θ_n, $n \geq 0$. For notational simplicity, we will often drop the subscript 0 and write $E_0 = E$ and $P_0 = P$. We shall use the following regularity conditions for proving the main results of the paper.

3.1 Conditions

We need to introduce some notation at this stage. For any two real numbers x, y, we write $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$. Let $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ denote the set of all integers. Also, let $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{Z}_+ = \{0, 1, \ldots\}$ respectively denote the set of all positive and the set of all nonnegative integers. For a positive definite matrix A of order $r \in \mathbb{N}$, we write Φ_A and ϕ_A to denote the distribution and the (Lebesgue) density of the $N(0, A)$ distribution in $\mathbb{R}^r$. For notational simplicity, we set $\Phi_{I_r} = \Phi$ and $\phi_{I_r} = \phi$ when $A = I_r$, the identity matrix of order r. For a function $f : \mathbb{R}^k \to \mathbb{R}$, we denote by $\partial^\nu f$ the partial derivative $\partial^{|\nu|}/(\partial t_1^{\nu_1} \ldots \partial t_k^{\nu_k}) f$ where $\nu \in \mathbb{N}^k$, $|\nu| = \sum_{i=1}^k \nu_i$ and $\nu! = \nu_1! \ldots \nu_k!$. When $|\nu| = 1$, we write ∂_i instead of ∂^ν to denote a partial derivative w.r.t. t_i. For a set B in $\mathbb{R}^r$, we write ∂B to denote its boundary and write B^ϵ for the set $B^\epsilon = \{x \in \mathbb{R}^r : \|x - y\| \leq \epsilon \quad \text{for some} \quad y \in B\}$. Also, let $e_1, \ldots, e_k$ denote the standard basis of unit vectors in $\mathbb{R}^k$. Let Θ_0 be an open neighborhood of θ_0.
The following are the regularity conditions on the log likelihood function l.

(A.1) For each ν, $1 \le |\nu| \le s+1$, $l(x;\theta)$ has a ν-th partial derivative $\partial^\nu l(x;\theta)$ with respect to θ on $\mathbb{R}^d \times \Theta$, and for $|\nu| \le s$, $\partial^\nu l(x;\theta)$ is continuous on Θ_0 for all $x \in \mathbb{R}^d$.

(A.2) There exists a constant $\delta \in (0,1)$ such that for all $n \in \mathbb{Z}_+$,

$$E_n\left[|\partial^\nu l(X_1;\theta_n)|^{(s+1)}\right] < \delta^{-1} \quad \text{for each} \quad \nu, 1 \le |\nu| \le s, \tag{3.2}$$

and

$$E_n\left[\sup_{|\theta-\theta_0|<\delta}\{|\partial^\nu l(X_1;\theta)|\}^s\right] < \delta^{-1} \quad \text{for each} \quad \nu, |\nu| = s+1. \tag{3.3}$$

(A.3) (i) For each $n \in \mathbb{Z}_+$, $E_n[\partial_i l(X_1;\theta_n)] = 0$ for $i = 1,\dots,k$.
(ii) The $k \times k$ matrices

$$I(\theta_n) = \{-E_n[\partial_i\partial_j l(X_1;\theta_n)]\}, \qquad D(\theta_n) = \{E_n[\partial_i l(X_1;\theta_n)\partial_j l(X_1;\theta_n)]\} \tag{3.4}$$

are non-singular and $I(\theta_n) = D(\theta_n)$ for all $n \in \mathbb{Z}_+$. Further, $\|I(\theta_n) - I(\theta_0)\| \to 0$ as $n \to \infty$.

For $n \in \mathbb{Z}_+$, define $Z_{in}^{[\nu]} = \partial^\nu l(X_i;\theta_n)$ and let $Z_{in} = (Z_{in}^{[\nu]})_{1\le|\nu|\le s}$ be the vector with coordinates indexed by the ν's. The dimension of Z_{in} is $m = \sum_{r=1}^{s}\binom{k+r-1}{r}$, and we arrange Z_{in} values such that the first k coordinates of Z_{in} are those with the indices $\nu = e_j$, $1 \le j \le k$. Some of the coordinates of Z_{in} may be linearly dependent. To deal with this, we suppose that there exist $m_0 \times m$ matrices A_n of rank $m_0 (\le m)$ such that the variables $\tilde{Z}_{in}$s defined by the relation

$$Z_{in} = \tilde{Z}_{in} A_n \tag{3.5}$$

are of dimension $m_0 \le m$ and are such that the coordinates of $\tilde{Z}_{1n}$ are linearly independent. We shall further suppose that the first m_1 of these coordinates are continuous variables and the remaining $m_2 = m_0 - m_1$ are lattice variables with minimal lattice $\mathbb{Z}^{m_2}$. We will write $\tilde{Z}_{in} = (\tilde{Z}_{in}^{(1)}, \tilde{Z}_{in}^{(2)})$, where $\tilde{Z}_{in}^{(1)}$ are the first m_1 coordinates and $\tilde{Z}_{in}^{(2)}$ are the last m_2 coordinates. For $\epsilon \in (0,\infty)$, define the set

$$\mathcal{C}(\epsilon) = \{(t,v) : t \in \mathbb{R}^{m_1}, v \in [-\pi,\pi]^{m_2}, \|t\| \wedge \|v\| \ge \epsilon\}.$$

We need the following additional set of conditions on the $\tilde{Z}_{in}$ vectors.

(A.4) (i) There exists a constant $\delta \in (0,1)$ such that for all $n \in \mathbb{Z}_+$,

$$E_n\left[\|\tilde{Z}_{1n}\|^{\max\{2s+1,m_1+1\}}\right] + E_n\left[\|\tilde{Z}_{1n}^{(2)}\|^{\max\{2s+1,m_1+1,m_2+1\}}\right] < \delta^{-1},$$

and the finite cumulants of the random vector Z_{1n} under θ_n converge to those of Z_{10} under θ_0 as $n \to \infty$.

(ii) For all $\varepsilon > 0$ there exists a $\delta \in (0,1)$ such that for all $n \in \mathbb{N}$,

$$\sup\left\{\left|E_n\left[\exp\left(it \cdot \tilde{Z}_{1n}^{(1)} + iv \cdot \tilde{Z}_{1n}^{(2)}\right)\right]\right| : (t,v) \in \mathcal{C}(\epsilon)\right\} \le 1 - \delta. \tag{3.6}$$

(A.5) (i) $\|A_n - A_0\| \to 0$ as $n \to \infty$. (ii) The $m_1 \times k$ matrix $A_0^{(11)}$ has full rank, where $A_0^{(11)}$ is the upper left hand corner of A_0.

(A.6) The $m_1 \times (k - k_1)$ matrix $(A_0^{(1)} I(\theta_0)^{-1/2})^{(12)}$ has full rank, where $A_0^{(1)}$ is the matrix consisting of the first k columns of A_0, the lower triangular matrix $I(\theta_0)^{-1/2}$ is the Cholesky factorization of $I(\theta_0)^{-1}$, and $(A_0^{(1)} I(\theta_0)^{-1/2})^{(12)}$ is the $m_1 \times (k - k_1)$ matrix of the first m_1 rows and columns $(k_1 + 1, \dots, k)$ of $A_0^{(1)} I(\theta_0)^{-1/2}$.

Relation (3.6) of Condition $(A.4)$ is called a uniform Cramer condition, and is required to establish an Edgeworth expansion for the continuous part $\tilde{Z}_{in}^{(1)}$, given the lattice part $\tilde{Z}_{in}^{(2)}$. Condition $(A.5)$ is used to assure that the part corresponding to the parameter $\theta^{(2)}$ in a first order Taylor approximation of the target statistic depends on the continuous part $\tilde{Z}_{in}^{(1)}$. Condition $(A.6)$ is used to assure the invariant property of the reparameterization. Note that in formulating the conditions, we include the limiting value θ_0 (i.e., the index $n = 0$) in all those conditions that ensure continuity of the resulting expansions at θ_0, e.g., (A.3) and (A.4)(i), and in conditions that simplify formulation of the uniformity conditions, e.g., conditions (A.5) and (A.6). Of all the conditions, the uniform version of the Cramer condition in (3.6) is perhaps the most difficult to verify. A simple sufficient condition for (3.6) that, in particular, allows one to dispense with the dependence of the condition on $\theta_n, n \geq 1$, is given by the following proposition.

Proposition 1 *Let $\{(X_n, Y_n)\}_{n \geq 0}$ be a collection of random vectors taking values in $I\!R^{m_1} \times Z\!\!Z^{m_2}$. Suppose that for each $n \geq 0$, the distribution of the random vector (X_n, Y_n) has an absolutely continuous component with respect to the product measure λ (say) of the Lebesgue measure on $I\!R^{m_1}$ and the counting measure on $Z\!\!Z^{m_2}$ with density $f_n(x, y)$. Assume that there exist a $c > 0$, a bounded open set $\mathcal{O} \equiv \prod_{j=1}^{m_1} (x_{j0} - a_j, x_{j0} + a_j) \subset I\!R^{m_1}$, and integers $l_1, l_2 \in Z\!\!Z$ such that with $B_0 = \mathcal{O} \times [l_1, l_2]^{m_2}$,*

(i) $\lim_{n \to \infty} f_n(x, y) = f_0(x, y)$ *for all* $(x, y) \in B_0$

(ii) $\lim_{n \to \infty} \int_{B_0} f_n d\lambda = \int_{B_0} f_0 d\lambda,$ *and*

(iii) $f_0(x, y) > c,$ for $(x, y) \in B_0$

Then, for any $\epsilon \in (0, \infty)$, there exists a $\delta = \delta(\epsilon) \in (0, 1)$ such that for $n = 0$ and for all $n \geq \delta^{-1}$,

$$\sup \{ |E[\exp(it \cdot X_n + iv \cdot Y_n)]| : (t, v) \in \mathcal{C}(\epsilon) \} \leq 1 - \delta.$$

Proposition 1 implies the inequality (3.6) for large values of n and this is adequate for the validity of the asymptotic results. In the next section, we describe the continuous Edgeworth expansion results for the likelihood-based statistics of Section 2.

3.2 Main Results

The first result concerns the MLE $\hat{\theta}_n$.

Theorem 1: *Assume that conditions (A.1)-(A.3) hold.*

(a) *Then there exist a sequence of statistics* $\{\hat{\theta}_n\}$ *and a constant* $a_1 \in (0, \infty)$, *independent of* n, *such that* $P_n(\| \hat{\theta}_n - \theta_n \| \leq a_1[\log n/n]^{1/2}, \hat{\theta}_n$ *solves (2.3))* $= 1 - o(n^{-(s-2)/2})$.

(b) *Assume that Conditions (A.4) and (A.5) hold. Then there exist polynomials* $q_j(\cdot;\theta)$ *(not depending on* n*) with coefficients that satisfy the continuity condition* $\lim_{n\to\infty} q_j(x;\theta_n) = q_j(x;\theta_0)$ *for all* $x \in I\!R^k$, *such that*

$$\sup_{B\in\mathcal{B}} \left| P_n(\sqrt{n}(\hat{\theta}_n - \theta_n) \in B) - \int_B \left[1 + \sum_{j=1}^{s-2} n^{-j/2} q_j(x, \theta_n) \right] \phi_{\Sigma_n}(x) dx \right| = o(n^{-(s-2)/2}),$$

where $\Sigma_n = I(\theta_n)^{-1}$ *and* $\mathcal{B}$ *is a collection of sets in* $I\!R^k$ *satisfying*

$$\sup_{B\in\mathcal{B}} \Phi_{\Sigma_n}([\partial B]^\delta) \leq C_1\delta, \quad \forall \delta \in (0,1), \quad n \geq \delta^{-1}, \tag{3.7}$$

where $C_1 \in (0, \infty)$ *is a constant.*

Theorem 1 extends Theorem 2.1 of Jensen (1993). Indeed, if we set $\theta_n \equiv \theta_0$ for all $n \geq 1$, then we get the Edgeworth expansion result of Jensen (1993) for the MLE over a larger class of Borel sets than the class of convex measurable sets in $I\!R^k$ considered in Jensen (1993). This follows from Corollary 3.2 of Bhattacharya and Ranga Rao (1986) (hereafter referred to as [BR]).

Next we consider the likelihood ratio statistic W_n and its Bartlett corrected version WB_n. Even in the presence of a discrete component, the Edgeworth expansions for smooth functions of $\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{Z}_{1n}$ are themselves smooth and do not involve the discontinuous see-saw functions that arise in the purely lattice case (cf. Theorem 23.1, [BR]). This is a consequence of a result of Göetz and Hipp (1978) on expansions for expectations of smooth function of means that are valid even for lattice variables. Although W_n admits an expansion in powers of n^{-1} in the complete data case (cf. Chandra and Ghosh (1979)), the same is not necessarily true for censored data. By integrating the expansion for the conditional probability of the continuous part with respect to the expansion for the discrete part, contributions from the series in powers of $n^{-1/2}$ enter into the Edgeworth expansions for W_n in the terms of order $O(n^{-3/2})$ and higher. As a result, we restrict attention only to a third order expansion for W_n and WB_n in the censored data case. This is adequate for investigating properties of the bootstrap approximation that we consider in the next section.

Let $A_n^{(12)}$ denote the $m_1 \times (k - k_1)$ submatrix of A_n consisting of the first m_1 rows and the last $(k - k_1)$ columns.

Theorem 2: *Suppose that Conditions (A.1)-(A.5) hold for* $s = 4$, $I(\theta_n)$ *is equal to the identity matrix and* $A_n^{(12)}$ *is of full rank.*

(a) *Let* $\tilde{R}_n = R_n(\theta_n; k-1)$ *where* $R_n(\cdot,\cdot)$ *is as defined in (2.8). Then,*

$$\sup_{x\in I\!R} \left| P_n(\tilde{R}_n \leq x) - \int_{-\infty}^x \left[1 + \sum_{r=1}^{2} n^{-r/2} \tilde{q}_{2+r}(u;\theta_n) \right] \phi(u) du \right| = o(n^{-1}).$$

(b) *There exists a polynomial* $\tilde{q}_j(\cdot;\theta)$*'s with coefficients that satisfy the continuity condition* $\lim_{n\to\infty} \tilde{q}_j(x;\theta_n) = q_j(x;\theta_0)$*, such that*

$$\sup_{0<u<\infty} \left| P_n(\tilde{W}_n \le u) - \int_0^u \left[1 + \frac{1}{n}\tilde{q}_1(v;\theta_n)\right] h_{k-k_1}(v)dv \right| = o(n^{-1})$$

where $\tilde{W}_n = W_n(\theta_n; k_1), W_n(\cdot,\cdot)$ *is as defined in (2.4) and where* h_{k-k_1} *is the (Lebesgue) density of the* χ^2*-distribution with* $(k-k_1)$ *degrees of freedom. If, in addition, the function* $B(\cdot)$ *in (2.6) is smooth in a neighborhood of* $E_{\theta_0}\bar{Z}_{0n}$*, then*

$$\sup_{0<u<\infty} \left| P_n(\tilde{WB}_n \le u) - \int_0^u \left[1 + \frac{1}{n}\tilde{q}_2(v;\theta_n)\right] h_{k-k_1}(v)dv \right| = o(n^{-1})$$

where $\tilde{WB}_n = WB_n(\theta_n; k_1)$ *and* $WB_n(\cdot,\cdot)$ *is as defined in (2.7).*

Note that part (b) only asserts that the Bartlett corrected version $\tilde{WB}_n$ has an error of approximation $O(n^{-1})$ by the limiting $\chi^2_{k-k_1}$ distribution for the censored case. It is not clear if $\tilde{q}_2(u;\theta_n) \equiv 0$ as in the complete data case, where the error of chi-squared approximation is known to be $O(n^{-2})$.

4 Results for Bootstrapped Statistics

In this section, we consider higher order accuracy of bootstrap approximations for the statistics considered in Section 2. All throughout, we suppose that the parametric bootstrap method is implemented by generating the bootstrap variables $X_1^*, \ldots, X_n^*$ from the estimated probability distribution $P_{\hat{\theta}_n}$ where $\hat{\theta}_n$ is the MLE of θ based on $X_1, \ldots, X_n$. Let θ_n^* denote the bootstrap version of the MLE, obtained by replacing $X_1, \ldots, X_n$ in the definition of $\hat{\theta}_n$ by $X_1^*, \ldots, X_n^*$. Also, recall that $I(\theta)$ denotes the Fisher information matrix of X_1 under θ. The Studentized version of $\hat{\theta}_n$ is given by

$$T_n = \sqrt{n}(\hat{\theta}_n - \theta_0) I(\hat{\theta}_n)^{1/2}$$

where $I(\theta)^{1/2}$ is a $k \times k$-matrix satisfying $I(\theta)^{1/2}[I(\theta)^{1/2}]' = I(\theta)$, obtained by the Cholesky decomposition of $I(\theta)$. Define the bootstrap version of T_n by

$$T_n^* = \sqrt{n}(\theta_n^* - \hat{\theta}_n) I(\theta_n^*)^{-1/2}.$$

Similarly, define the bootstrap versions R_n^*, W_n^* and WB_n^* of R_n, W_n and WB_n by replacing $X_1, \ldots, X_n$ by $X_1^*, \ldots, X_n^*$ and θ_0 by $\hat{\theta}_n$ in (2.8), (2.4) and (2.7), respectively. For proving the results, we shall suppose that for each $\theta \in \Theta_0$, the random vector $Z_1(\theta) \equiv (\partial^\nu l(X_1;\theta))_{1\le|\nu|\le s}$ can be transformed to an m-dimensional vector $\tilde{Z}_1(\theta)$ as in (3.5) for some $m_0 \times m$ matrix $A(\theta)$ such that the first m_1 components $\tilde{Z}_1^{(1)}(\theta)$ of $\tilde{Z}_1(\theta)$ take values in $\mathbb{R}^{m_1}$ and the last m_2 components $\tilde{Z}_1^{(2)}(\theta)$ of $\tilde{Z}_1(\theta)$ are discrete with minimal lattice $\mathbb{Z}^{m_2}$. Further, the distribution of $\tilde{Z}_1(\theta)$ under θ has an absolutely continuous component w.r.t λ with density $f(x,y;\theta), x \in \mathbb{R}^{m_1}, y \in \mathbb{Z}^{m_2}, \theta \in \Theta_0$. This would allow us to verify the uniform Cramer condition (A.4)(ii) along different realizations of the sequence $\{\hat{\theta}_n\}$ that lie in a set of probability 1 under θ_0.

Next, write P_* for the conditional probability under $\hat{\theta}_n$, given $X_1, X_2, \ldots, X_n$. Then we have the following result.

We shall use the following modified versions of some of the Conditions (A.1)-(A.6). Recall that we set $E_{\theta_0} = E_0 = E$ for notational simplicity.

(A.2)$'$ There exists $\delta \in (0,1)$ such that

(i) $E|\partial^\nu l(X_1, \theta_0)|^{s+1} < \delta^{-1}$ and the cumulants of $Z_1(\theta)$ under θ up to order $(s+1)$ are continuous over Θ_0.

(ii) $E_\theta\{\sup_{\|\theta-\theta_0\|\leq\delta} |\partial^\nu l(X_1;\theta) \|^s\} < \delta^{-1}$ for all $\theta \in \Theta_0$.

(A.3)$'$ (i) $E_\theta \partial_i l(X_1, \theta) = 0$ for all $\theta \in \Theta_0, 1 \leq i \leq k$.

(ii) $I(\theta_0)$ of (3.4) is nonsingular and $I(\theta) = D(\theta)$ for all $\theta \in \Theta_0$.

(A.4)$'$ (i) $E_\theta[\| \tilde{Z}_1^{(1)}(\theta) \|^{\max\{2s+1, m_1+1\}}] + E_\theta[\| Z_1^{(2)}(\theta) \|^{\max\{2s+1, m_1+1, m_2+1\}}] < \infty$ for all $\theta \in \Theta_0$ and all finite cumulants of the random vector $\tilde{Z}_1(\theta)$ under θ are continuous on Θ_0.

(ii) There exists a $c \in (0,\infty)$ such that $f(x,y;\theta_0) > c$ for all $(x,y) \in B_0$ and the function $g(\theta; B_0) \equiv \int_{B_0} f(x,y,\theta) d\lambda(x,y)$ is continuous at $\theta = \theta_0$, where B_0 is as defined in the statement of Proposition 1.

(A.5)$'$ (i) $A(\theta)$ is continuous at $\theta = \theta_0$.

(ii) The matrix $A_0^{(11)}$ of Condition (A.5) is of full rank.

Condition (A.2)$'$-(A.5)$'$ are stronger versions of (A.2)-(A.5) and ensure that (A.2)-(A.5) holds for every sequence $\{\theta_n\}$ that converges to θ_0. Conditions (A.1) and (A.6) did not involve the sequence $\{\theta_n\}$ and therefore, may be used in this section without further modifications. For notational simplicity, we set (A.1)$'$=(A.1), (A.6)$'$=(A.6). Then, we have the following results.

Theorem 3: *Suppose that Conditions (A.1)$'$-(A.5)$'$ hold with $s = 3$.*

(a) Then,

$$\sup_{B\in\mathcal{B}} |P_*(T_n^* \in B) - P(T_n \in B)| = o(n^{-1/2})\, a.s.(P).$$

(b) If, in addition, $s \geq 4, I(\theta_0) = I_k$ and $A_0^{(12)}$ is of full rank and Condition (A.6) holds, then

$$\sup_{u\in\mathbb{R}} |P_*(R_n^* \leq u) - P(R_n \leq u)| = o(n^{-1/2}) \quad a.s.(P);$$
$$\sup_{0<u<\infty} |P_*(W_n^* \leq u) - P(W_n \leq u)| = o(n^{-1}) \quad a.s.(P);$$
$$\sup_{0<u<\infty} |P_*(WB_n^* \leq u) - P(WB_n \leq u)| = o(n^{-1}) \quad a.s.(P).$$

Thus, it follows that the bootstrap improves upon the normal approximation to the distribution of T_n and is second order correct even in presence of censoring. If we assume that $s \geq 4$, then the $o(n^{-1/2})$ term is indeed $O(n^{-1})$ in P_{θ_0}-probability. Part (b) shows that similar improvements over the limiting normal and χ^2-approximations are achieved by the bootstrap for the SRLLR statistic R_n and the likelihood ratio statistic W_n, respectively. The condition $I(\theta_0) = I_k$ is not a strong restriction, as it can be established through a reparametrization (but the regularity conditions (A.5)′-(A.6)′ would have to be restated accordingly).

5 Numerical Results

The theoretical results in this paper hold under standard regularity conditions. As shown in the apendix, those conditions hold, for example, for the smallest extreme value, normal and logistic distributions location-scale distributions. The results are also valid for the corresponding log-location-scale distributions (i.e., the two-parameter Weibull, lognormal, and loglogistic distributions).

To explore the finite sample performance of the asymptotic results in Sections 3 and 4, we conducted a simulation study using the two-parameter Weibull distribution model with Type I censored data. We also incorporate the complete data case in the simulation study to gain some insight on the effects of censoring on accuracy of the likelihood-based methods for one- and two-sided confidence intervals. In Section 5.1, we describe the two parameter Weibull distribution model and in Section 5.2, we describe the simulation design and relevant formulas of the confidence intervals. We present the results of the Weibull simulation study in Section 5.3. Results for the other loglocation scale distributions show distributions with smaller variances.

5.1 The Two Parameter Log-Location-Scale Distribution Model

for example, the logarithm of a log-location scale random variable has the corresponding location-scale distribution. For examle, the logarithm of a Weibull random variable has a smallest extreme value distribution. Suppose that the continuous random variable $X = \log(T)$ has density $\phi_{\mathrm{LS}}[(x-\mu)/\sigma]/\sigma$ and cdf $\Phi_{\mathrm{LS}}[(x-\mu)/\sigma]$, where $(\mu, \sigma) = \theta$ is the unknown parameter in an open set $\Theta \subset I\!R^2$. Let t_c denote the censoring time and define $\delta = 1$ for a failure and $\delta = 0$ for a censored observation. The observations are $x_1 = \log(t_1), \ldots, x_n = \log(t_n)$. Let $x_c = \log(t_c)$. The log likelihood of an observation x_i is

$$l(x_i;\theta) = \delta_i \left\{ -\log(\sigma) + \log\left[\phi_{\mathrm{LS}}\left(\frac{x_i-\mu}{\sigma}\right)\right]\right\} + (1-\delta_i)\log\left[1-\Phi_{\mathrm{LS}}\left(\frac{x_c-\mu}{\sigma}\right)\right]. \quad (5.8)$$

One might be interested in the location or the scale parameter or in a particular quantile or other function of these parameters. We do the development for estimating a particular quantile. Other functions of the parameters can be obtained analogously. Let x_p be the p quantile of the distribution $\Phi_{\mathrm{LS}}[(x-\mu)/\sigma]$, and $u_p = \Phi_{\mathrm{LS}}^{-1}(p)$. Then $x_p = \mu + u_p\sigma$ and $t_p = \exp(x_p)$ is the p quantile of the Weibull distribution. The confidence intervals (bounds) for t_p can be obtained by taking the antilog of transformation of the confidence intervals (bounds) for x_p. The likelihood in (5.8) can be rewritten as a function of (σ, x_p)

$$l(x_i;(\sigma,x_p)) = \delta_i \left\{ -\log(\sigma) + \log\left[\phi_{\mathrm{LS}}\left(\frac{x_i - x_p}{\sigma} - u_p\right)\right]\right\} + \\ (1-\delta_i)\log\left[1 - \Phi_{\mathrm{LS}}\left(\frac{x_c - x_p}{\sigma} - u_p\right)\right]. \tag{5.9}$$

With l smooth enough and ϕ having light tails, it can be shown that Conditions (A.1)$'$ -(A.3)$'$, stated in Section 4 are satisfied. See the Appendix for details. Then for $|\nu| \le 4$,

$$Z_i = \left(\frac{\partial l(X_i;(\sigma,x_p))}{\partial x_p}, \frac{\partial l(X_i;(\sigma,x_p))}{\partial \sigma}, \frac{\partial^2 l(X_i;(\sigma,x_p))}{\partial x_p^2}, \frac{\partial^2 l(X_i;(\sigma,x_p))}{\partial x_p \partial \sigma}, \right. \\ \left. \ldots, \frac{\partial^4 l(X_i;(\sigma,x_p))}{\partial \sigma^4}\right),$$

where Z_i is a 14 dimensional vector. Transform Z_i into a $m_0 = m_1 + m_2$ dimensional vector $\tilde{Z}_i$ with linearly independent coordinates for which the first m_1 ordinates are continuous and last m_2 coordinates are discrete. The form of $\tilde{Z}_i$ depends on the distribution of the observations. For the SEV, normal, and logistic distributions, $\tilde{Z}_i$ is shown in the Appendix. Note that δ_i is the only discrete part of Z_i, so it is the only discrete part of $\tilde{Z}_i$. By Proposition 1, Condition (A.4)$'$ is satisfied here.

The first two elements of Z_i are linearly independent when data come from the SEV, normal or logistic distribution (see the Appendix). The first two elements of the first two columns of $A^{(11)}$ are $(1, 0)$ and (c_1, c_2) respectively, where c_1, c_2 are non-zero constants (that could depend on the parameters), hence $A^{(11)}$ has full rank 2. For the SEV, normal, and logistic distributions, (c_1, c_2) is just $(0, 1)$. So Condition (A.5)$'$ holds.

Because the first m_1 rows of $A^{(1)}$ gives $A^{(11)}$ as described in Section 3.1 and $I(\theta_0)^{-1/2}$ is a lower triangular positive definite matrix, $(A^{(1)} I(\theta_0)^{-1/2})^{(11)}$ is a m_1 dimensional vector that has rank 1. Thus, Condition (A.6)$'$ holds. Theorems 1-3 tell us that the procedure based on the bootstrap log likelihood ratio statistic or its corresponding signed square root procedures can be used to construct two-sided (one-sided) confidence intervals (bounds) that are second order accurate.

5.2 The Simulation Design

5.2.1 Confidence Intervals

This section briefly describes the different particular confidence interval procedures that we consider in our simulation study. For more details, see the given references.

Log LR method (LLR). The distribution of W is approximately χ^2_1. Thus an approximate $100(1-\alpha)\%$ confidence interval can be calculated from $\min\{W^{-1}(\chi^2_{(1-\alpha,1)})\}$ and $\max\{W^{-1}(\chi^2_{(1-\alpha,1)})\}$, where $W^{-1}[\cdot]$ is the inverse mapping and $\chi^2_{(1-\alpha,1)}$ is the $1-\alpha$ quantile of χ^2 distribution with 1 degree of freedom.

Log LR Bartlett corrected method (LLRB). Let $WB = W/\mathrm{E}(W)$. In general one must substitute an estimate for $\mathrm{E}(W)$ computed from one's data. For complicated problems (e.g., those involving censoring) it is necessary to estimate of $\mathrm{E}(W)$ by using simulation. Then an approximate $100(1-\alpha)\%$ confidence interval can be obtained by using $\min\{WB^{-1}[\chi^2_{(1-\alpha,1)}]\}$ and $\max\{WB^{-1}[\chi^2_{(1-\alpha,1)}]\}$.

Parametric transformed bootstrap-t method (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 of the ML estimator. Let $z_{g(\widehat{\theta_1}^*)_{(\alpha)}}$ be the α quantile of the distribution of $[g(\widehat{\theta}_1^*) - g(\widehat{\theta}_1)]/\widehat{\mathrm{se}}^*[g(\widehat{\theta}_1)]$, where $\widehat{\mathrm{se}}^*[g(\widehat{\theta}_1)]$ is the bootstrap version of $\widehat{\mathrm{se}}[g(\widehat{\theta}_1)]$. We choose $\widehat{\mathrm{se}}[g(\widehat{\theta}_1)]$ to be $g'(\widehat{\theta}_1)\widehat{I}_{\widehat{\boldsymbol{\theta}}}^{(1,1)}g'(\widehat{\theta}_1)$, where $\widehat{I}_{\widehat{\boldsymbol{\theta}}}$ is the local estimate of $I_{\boldsymbol{\theta}}$. For estimating quantiles of a positive random variable we take g to be the log transformation. An approximate $100(1-\alpha)\%$ confidence interval for θ_1 can be computed from $g^{-1}\{g(\widehat{\theta}_1) - z_{g(\widehat{\theta_1}^*)_{(1-\alpha/2)}}\widehat{\mathrm{se}}[g(\widehat{\theta}_1)]\}$ and $g^{-1}\{g(\widehat{\theta}_1) - z_{g(\widehat{\theta_1}^*)_{(\alpha/2)}}\widehat{\mathrm{se}}[g(\widehat{\theta}_1)]\}$.

Parametric bootstrap bias-corrected accelerated method (PBBCA). Efron and Tibshirani (1993, Section 14.3) showed an easy way to obtain BC$_a$ confidence intervals. An approximate $100(1-\alpha)\%$ confidence interval is given by $(\widehat{\theta_1}^*_{(\alpha_1)}, \widehat{\theta_1}^*_{(\alpha_2)})$. Where $\widehat{\theta_1}^*_{(\alpha)}$ is the α quantile of the distribution of $\widehat{\theta}_1^*$ and

$$\alpha_1 = \Phi\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z_{\alpha/2}}{1-\widehat{a}(\widehat{z}_0 + z_{\alpha/2})}\right), \quad \alpha_2 = \Phi\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z_{1-\alpha/2}}{1-\widehat{a}(\widehat{z}_0 + z_{1-\alpha/2})}\right),$$
$$\widehat{z}_0 = \Phi^{-1}\left(\frac{\#\{\widehat{\theta}_1^*(b) < \widehat{\theta}_1\}}{B}\right), \quad \widehat{a} = \frac{\sum_{i=1}^n \left(\widehat{\theta}_{1[\cdot]} - \widehat{\theta}_{1[i]}\right)^3}{6\left[\sum_{i=1}^n \left(\widehat{\theta}_{1[\cdot]} - \widehat{\theta}_{1[i]}\right)^2\right]^{3/2}}.$$

Usually Φ is taken to be the standard normal cdf. Here $\widehat{\theta}_{1[i]} = \widehat{\theta}_1(X_{[i]}), X_{[i]}$ is the original sample with the ith point x_i deleted, $\widehat{\theta}_{1[\cdot]} = \sum_{i=1}^n \widehat{\theta}_{1[i]}/n$, z_α is the α quantile of normal distribution, and B is the number of the bootstrap samples, and $\widehat{\theta}_1^*(b), b = 1, ..., B$ are bootstrap versions of $\widehat{\theta}_1$.

If there is an increasing function ψ_n (the exact form need not be known) such that

$$\Pr\left\{\frac{\psi_n(\widehat{\theta}_1) - \psi_n(\theta_1)}{1 + a\psi_n(\theta_1)} + z_0 \le x\right\} = \Phi(x),$$

then the BC$_a$ CI procedure is exact.

Parametric bootstrap signed square root LLR method (PBSRLLR). Suppose that $r_{\widehat{\theta}_1{}^*_{(\alpha)}}$ is the α quantile of the bootstrap distribution of a SRLLR statistic, $R(\theta_1)$. Then an approximate $100(1-\alpha)\%$ confidence interval can be computed from
$\min\{R^{-1}(r_{\widehat{\theta}_1{}^*_{(\alpha/2)}}), R^{-1}(r_{\widehat{\theta}_1{}^*_{(1-\alpha/2)}})\}$ and $\max\{R^{-1}(r_{\widehat{\theta}_1{}^*_{(\alpha/2)}}), R^{-1}(r_{\widehat{\theta}_1{}^*_{(1-\alpha/2)}})\}$.

For easy reference later on, Table 1 summarizes the abbreviations used for different confidence interval procedures.

Table 1: Abbreviations of the methods in simulation study

LLR	Log likelihood ratio
LLRB	Log likelihood ratio Bartlett corrected
PTBT	Parametric transformed bootstrap-t
PBBCA	Parametric bootstrap bias-corrected accelerated
PBSRLLR	Parametric bootstrap signed squared root LLR

5.2.2 The Simulation Set Up

Let T be a random variable having a Weibull distribution, then $X = \log(T)$ has a smallest extreme value (SEV) distribution with density $\phi_{SEV}(z)/\sigma$ and cdf $\Phi_{SEV}(z)$, where $\phi_{SEV}(z) = \exp[-z - \exp(z)]$, $\Phi_{SEV}(z) = 1 - \exp[-\exp(z)]$ and $z = (x - \mu)/\sigma$, $-\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0$. Our simulation was designed to study the following experimental factors:

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

We used 5000 Monte Carlo samples for each p_f and $\mathrm{E}(r)$ combination. The number of bootstrap replications was $B = 10000$. The levels of the experimental factors used were $p_f = .01$, .1, .3, .5, .9, 1 and $\mathrm{E}(r) = 3$, 5, 7, 10, 15 and 20. For each Monte Carlo sample we obtained the ML estimates of the scale parameter and the quantiles $\log(t_p)$, $p = .01$, .05, .1, .3, .5, .632 and .9, where $\mu \cong \log(t_{.632})$. The one-sided $100(1-\alpha)\%$ confidence bounds (CBs) were calculated for α =.025 and .05. Hence the 90% and 95% two-sided CIs can be obtained by combining the upper and lower CBs. Without loss of generality, we sampled from an SEV distribution with $\mu = 0$ and $\sigma = 1$.

Because the number of failures before the censoring time t_c is random, it is possible to have as few as $r = 0$ or 1 failures in the simulation, especially when $\mathrm{E}(r)$ is small. The PBBCA procedure requires at least $r = 2$ failures before the censoring time in order to estimate the accelerated constant. Therefore, we calculate the results conditionally on the cases with $r > 1$.

Let $1 - \alpha$ be the nominal (user-specified) coverage probability (CP) of a procedure for constructing a confidence interval, and let $1 - \breve{\alpha}$ denote the corresponding Monte Carlo evaluation of the actual coverage probability $1-\alpha^{'}$. The standard error of $\breve{\alpha}$ is approximately $\mathrm{se}(1 - \breve{\alpha}) = [\alpha^{'}(1 - \alpha^{'})/n_s]^{1/2}$, where n_s is the number of Monte Carlo simulation trials. For a 95% confidence interval from 5000 simulations the standard error of the CP estimation is $[.05(1 - .95)/5000]^{1/2} = .0031$ if the procedure is correct. The Monte Carlo error is approximately $\pm 1\%$. We say the procedure or the method for the 95% confidence region is adequate if the Monte Carlo evaluation of CP is within $\pm 1\%$ error of the nominal CP.

5.3 Simulation Results

In this section, we present some of the major findings from our simulation.

Figure 1 shows the coverage probability of the procedures for the one-sided approximate 95% CBs for the parameter σ from the seven methods for five different proportion failing values. Figure 2 is the same type of graph as Figure 1 for $t_{.1}$, the .1 quantile of Weibull distribution. Figure 3 shows CPs of these procedures when $p_f = .5$ for different quantiles. Figures 4 and 5 and some other similar graphs, which are not shown here, present a closer comparison of CP for methods and parameters. Figure 6 and Figure 7 show the coverage probability of these procedures for 90% two-sided confidence intervals. We summarize the simulation results briefly as follows:

- Using a Bartlett correction for the LLR method does not improve the coverage probability of the procedure for one-sided confidence bounds. For one-sided confidence bounds, the LLR and LLRB methods are adequate when the expected number of failures ≥ 20. For two-sided confidence intervals, the LLR method is adequate when the expected number of failures is more than 15 and the LLR method with a Bartlett correction is very accurate even for an expected number of failures as small as 7.

- The bootstrap-t method is an accurate procedure for the scale parameter. When the quantity of interest is t_p, the p quantile, where p is close to the proportion failing, the one-sided lower confidence bound procedure is anti-conservative. The bootstrap-t method gives accurate coverage probabilities for all functions of the parameters when the number of failures exceeds 20. This is because the distribution of $\widehat{t_p}$ is approximately discrete.

- The BC$_a$ method for both one-sided confidence bounds and two-sided confidence intervals is adequate when the number of failures exceeds 20.

- The PBSRLLR method for the one-sided confidence bounds and two-sided confidence intervals is adequate except when the number of failures is less than 15 and the quantity of interest is the p quantile where p is close to the proportion failing.

For Type I censored data, we can draw the following conclusion. If our interest is in constructing one-sided confidence bounds, the PBSRLLR method provides better coverage probability with a small expected number of failures (like 10). For two-sided confidence intervals, the PBSRLLR and LLRB methods provide accurate procedures. The LLRB method gives more accurate results even when the expected number of failures is as small as 7. The two-sided confidence interval from the PBSRLLR is more symmetric than that from other methods in the sense that the confidence level of one side of the interval is close to the confidence level of the other side of the interval.

6 Summary of Results and Possible Areas for Future Research

In this paper we prove that the distributions of likelihood ratio statistics and their signed square root can be approximated by their bootstrap distribution up to the second order when the underlying sampling distribution is partly discrete. This result can be applied to

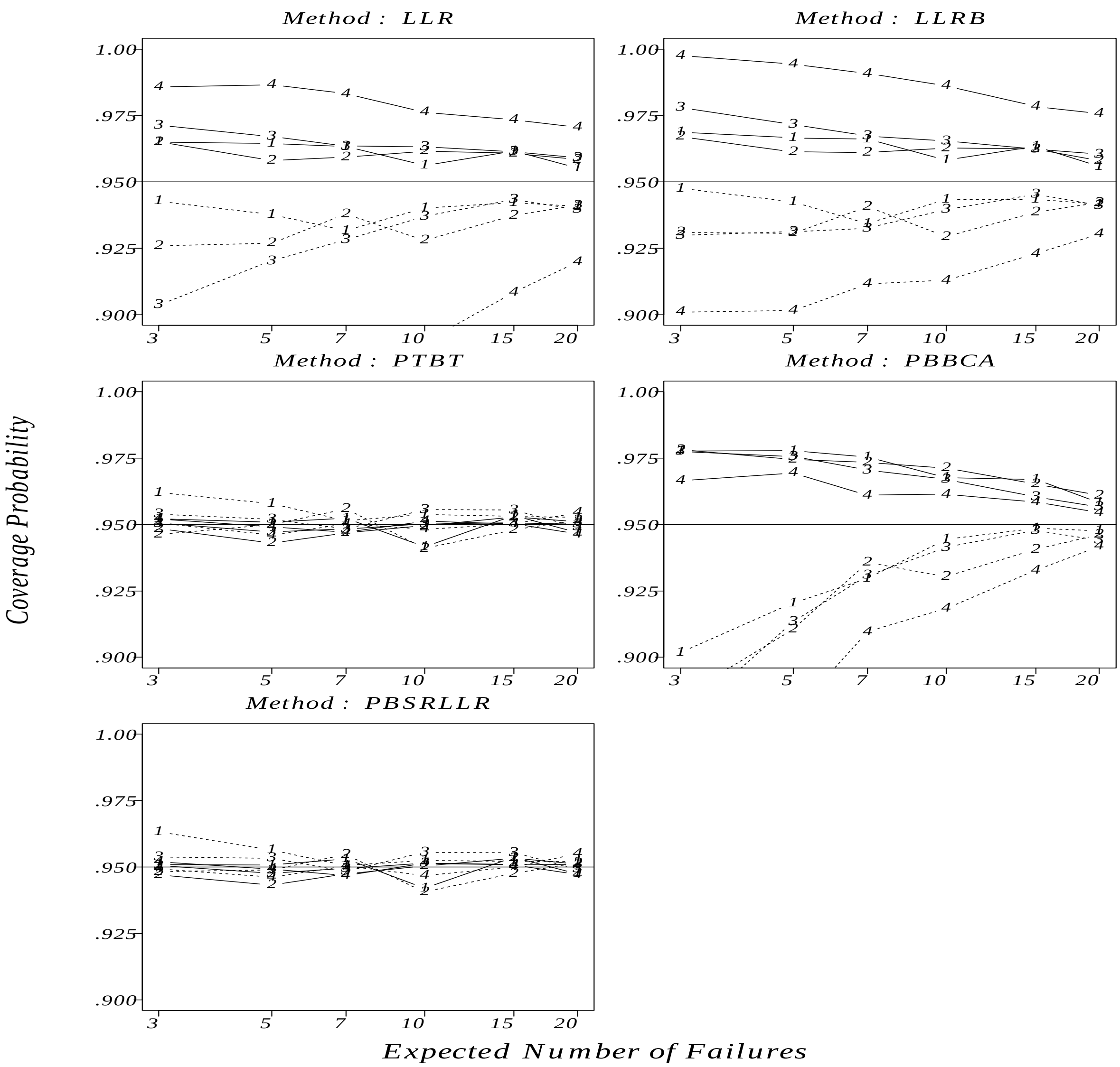

Figure 1: Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for parameter σ. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (.01, .1, .3, .5, 1), respectively. Dotted and solid lines correspond to upper and lower bounds,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

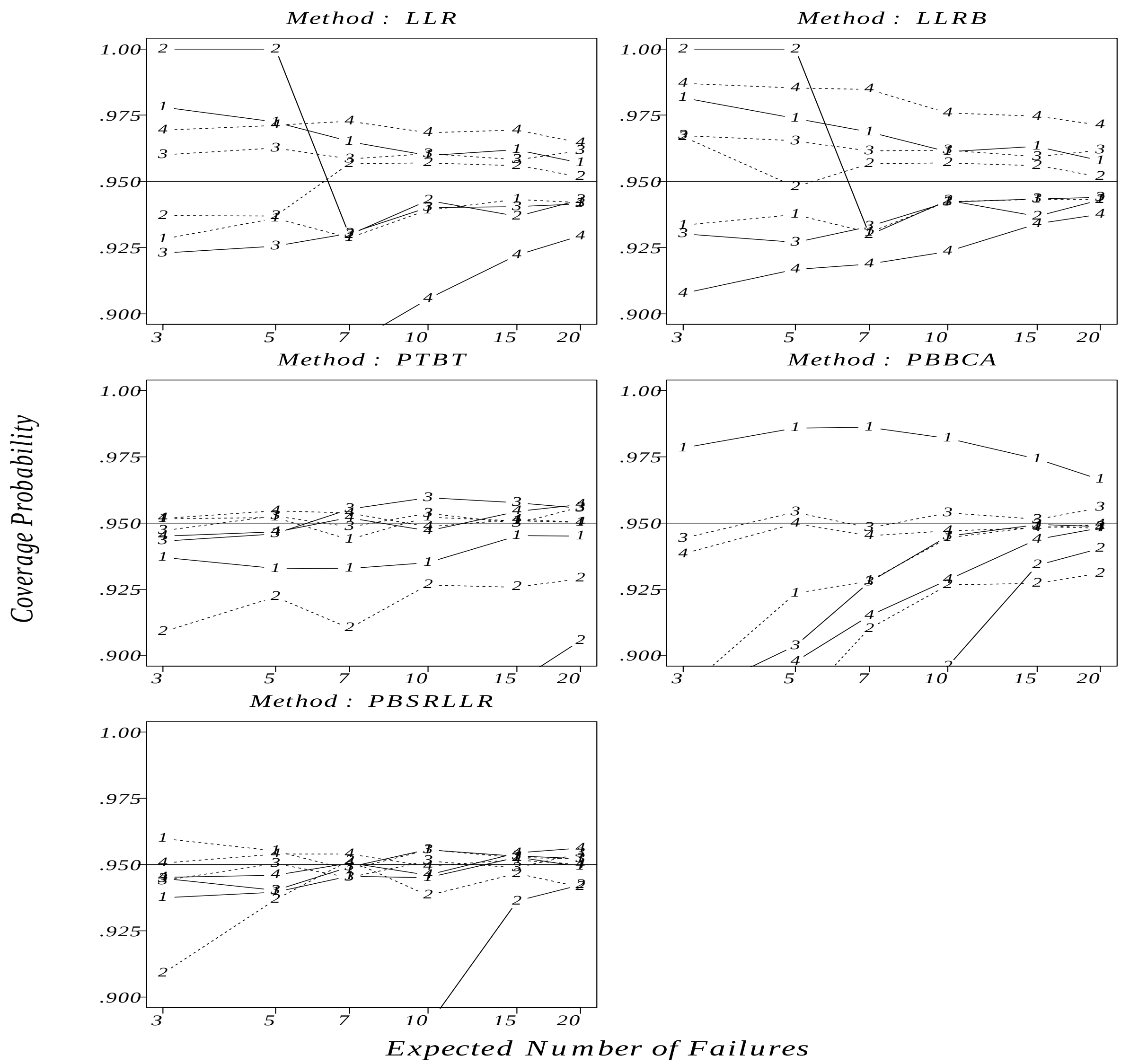

Figure 2: Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (.01, .1, .3, .5, 1), respectively. Dotted and solid lines correspond to upper and lower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

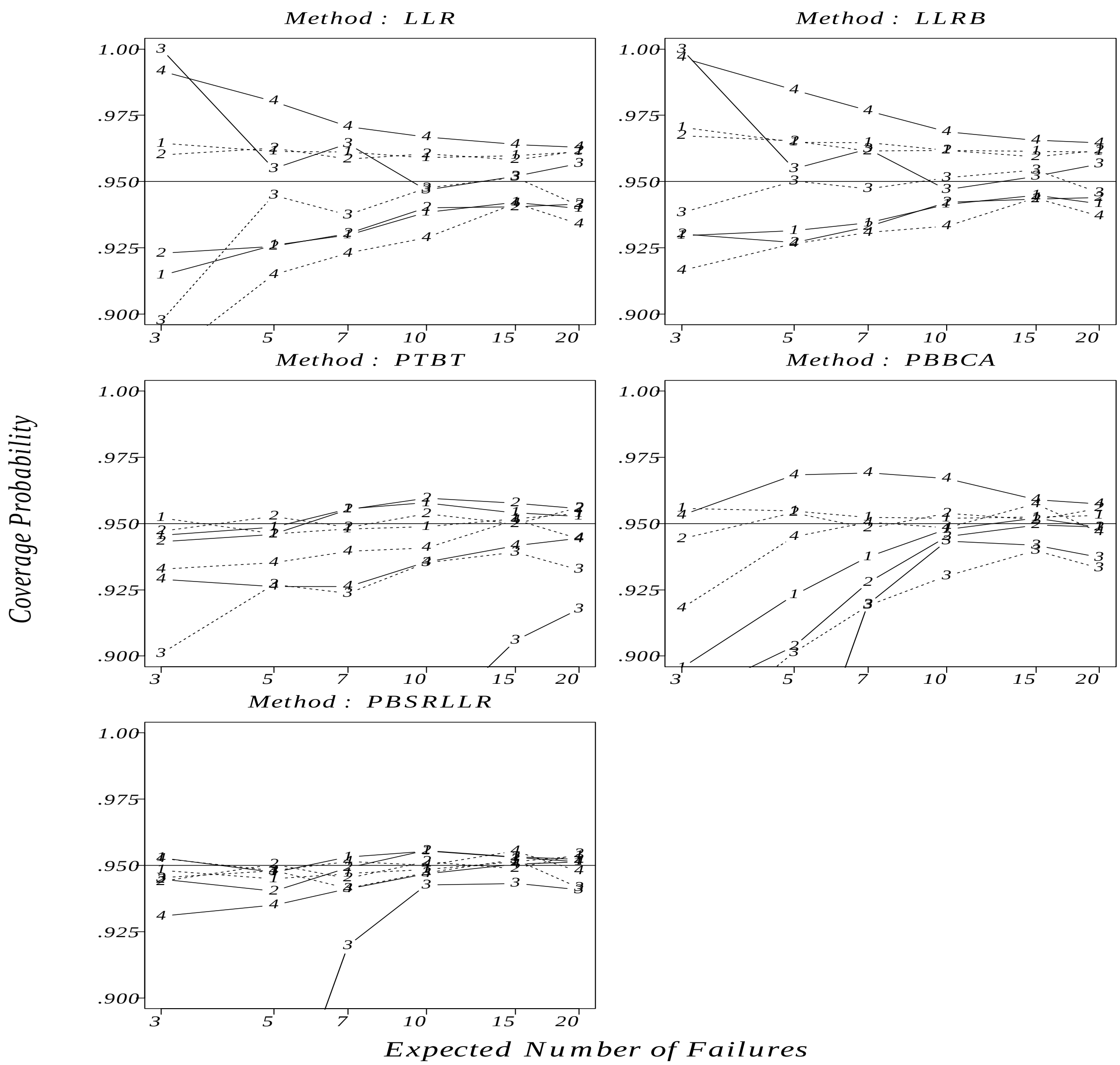

Figure 3: Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for proportion failing $p_f = .5$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to t_p's, $p = (.01, .1, .5 ,.632, 9)$, respectively. Dotted and solid lines correspond to upper and lower bounds, respectively.

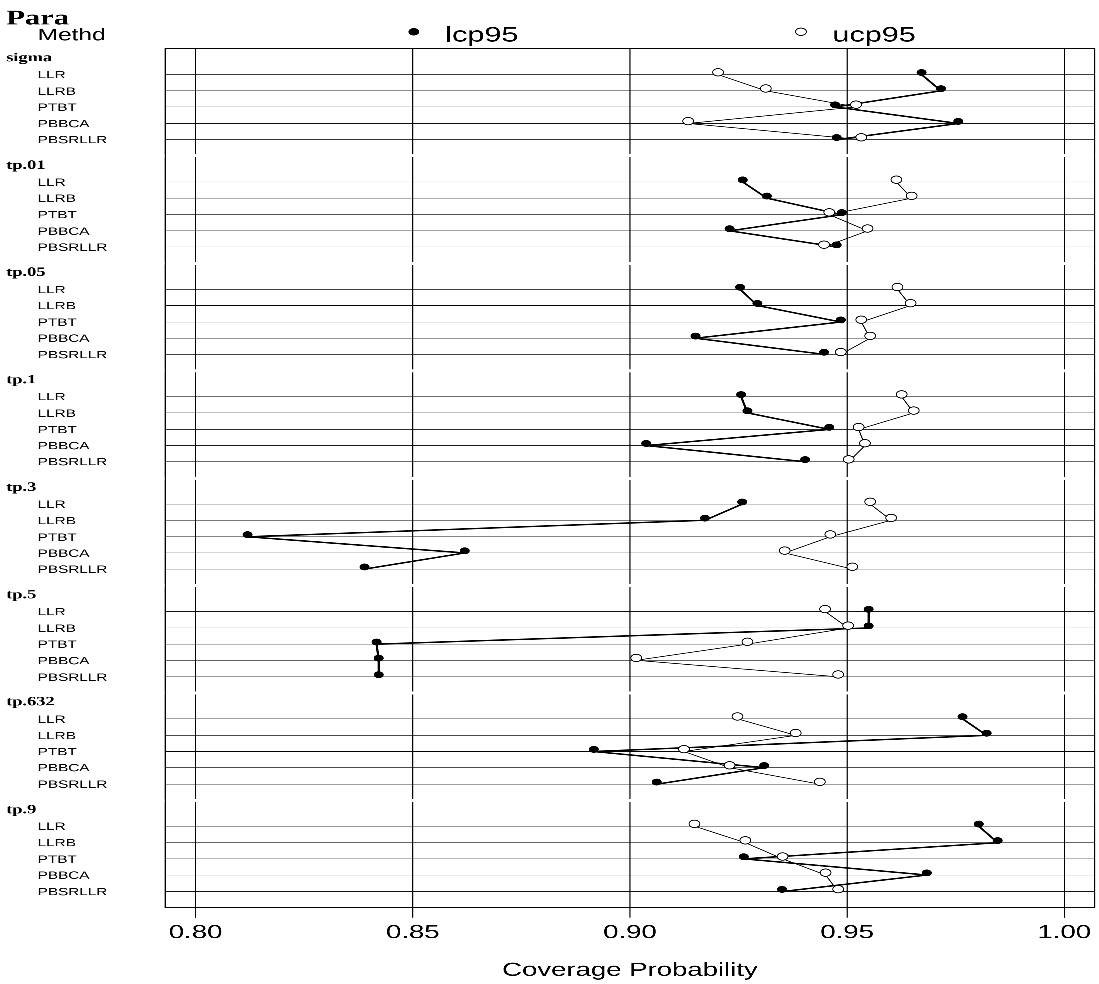

Figure 4: Coverage probability plot for approximate 95% one-sided confidence interval procedures in the case $\mathrm{E}(r) = 5$ and $p_f = .5$.

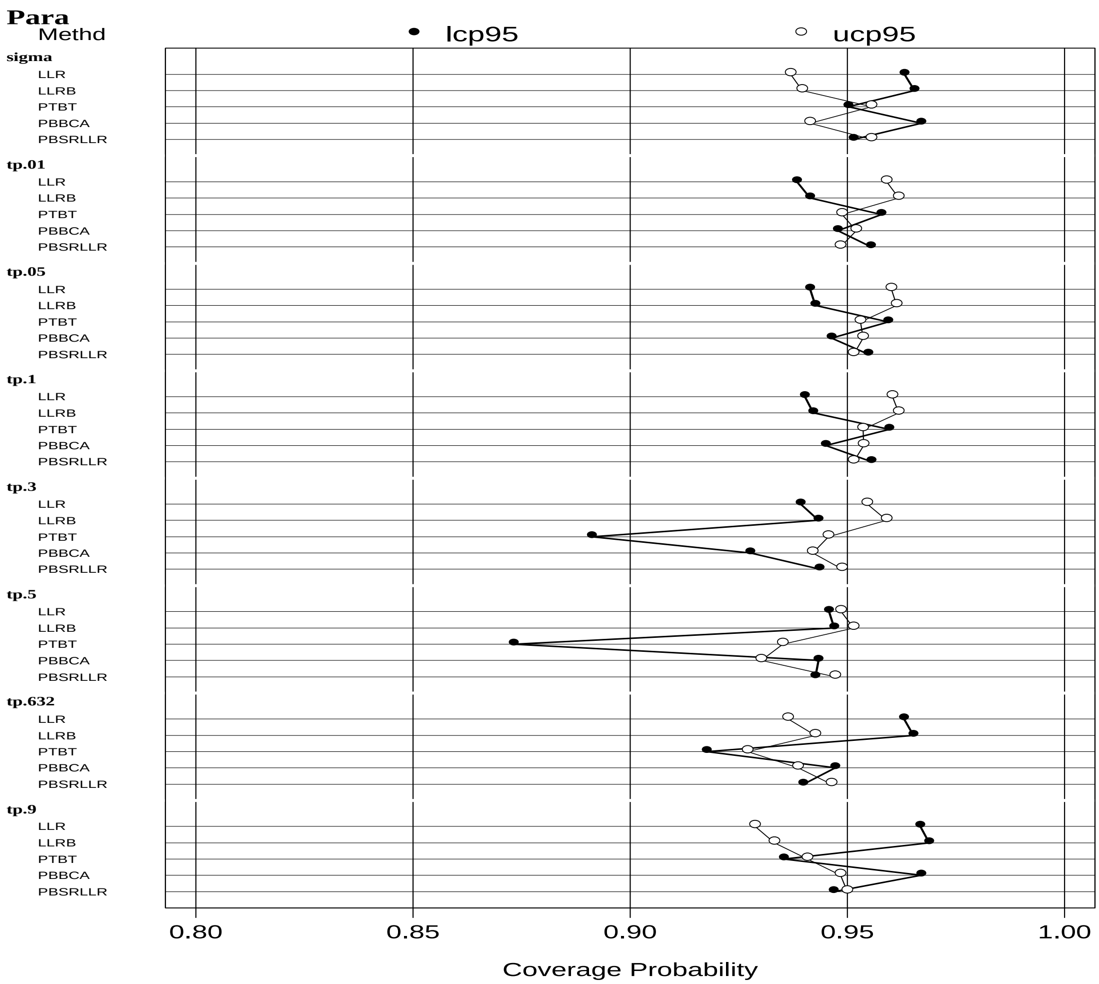

Figure 5: Coverage probability plot for approximate 95% one-sided confidence interval procedures in the case $\mathrm{E}(r) = 10$ and $p_f = .5$.

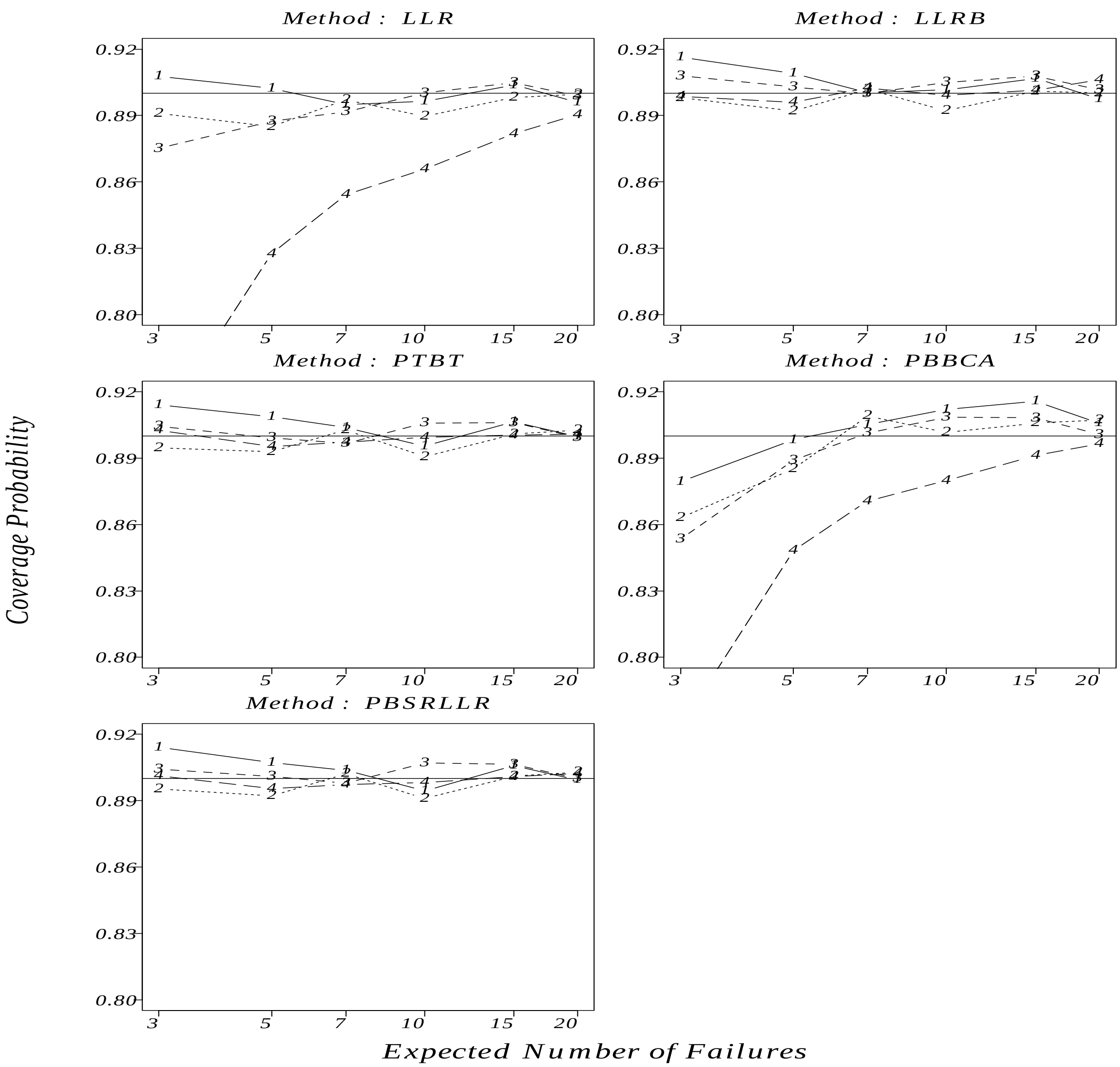

Figure 6: Coverage probability versus expected number of failures plot for two-sided approximate 90% CI procedures for parameter σ. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (.01, .1, .3, .5, 1), respectively. Dotted and solid lines correspond to upper and lower bounds, respectively.

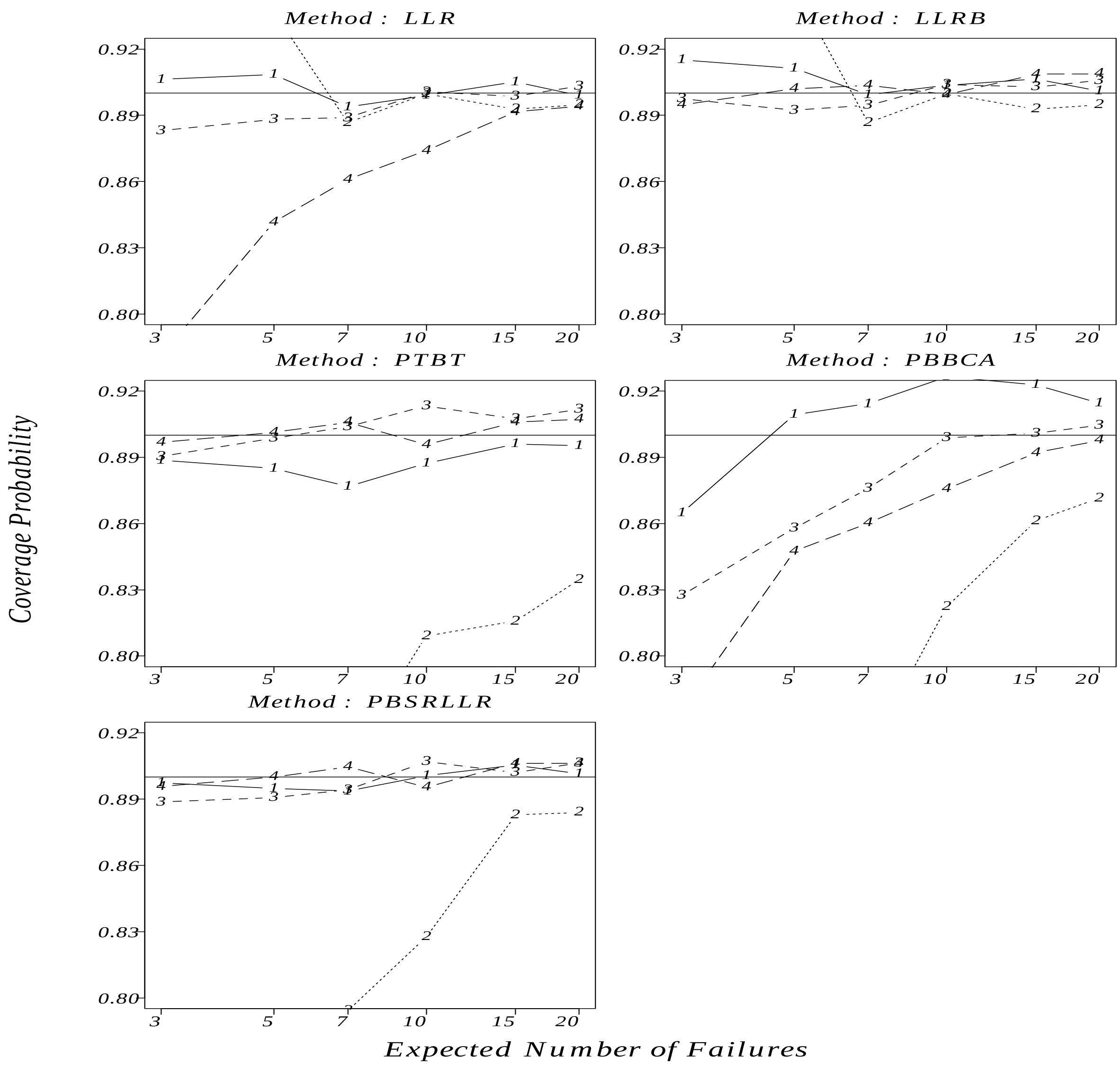

Figure 7: Coverage probability versus expected number of failures plot for two-sided approximate 90% CI procedures for parameter $t_{.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (.01, .1, .3, .5, 1), respectively. Dotted and solid lines correspond to upper and lower bounds, respectively.

investigate accuracy of bootstrap procedures for constructing one-sided confidence bounds, two-sided confidence intervals or joint confidence region for complete or censored data.

Examples like the one-parameter exponential model with Type I censoring and logistic regression given by Jensen (1989, 1993) illustrate some applications. For the two-parameter Weibull distribution model when data are Type I censored, our simulation study compares several commonly suggested methods (Bootstrap-t and BC_a) and our theorems of likelihood ratio statistics calibrated by bootstrap procedures. The simulation provides a clear view of the small sample properties of these statistics.

We can draw the following conclusions from our simulations involving Type I censored data. If one-sided confidence bounds are of interest, the PBSRLLR method provides better coverage probability when the expected number of failures exceeds 10. If two-sided confidence intervals are of interest, the PBSRLLR and LLRB methods provide accurate procedures and moreover, the LLRB methods give accurate coverage probability when the expected number of failures exceeds 7. Although the LLRB method for two-sided confidence intervals is the most accurate one in coverage probability among these methods, the resulting two-sided confidence interval is not symmetric in the sense that the confidence level of one side of the interval is larger than the nominal confidence level and the confidence level of the other side of the interval is smaller than the nominal one.

Some possible areas for further research are:

- Our examples show that the theorems in Sections 3 and 4 can be applied to the location-scale model with Type I censoring data. For other kinds of censoring and distributions, Conditions (A) and (A)$'$ can be expected to hold when the model distributions are smooth and without overly heavy tails. It would be of interest to study the finite sample coverage probabilities for such distributions.

- Although the order of accuracy is the same for different parameters of interest in the theorem, our simulation study shows that, in small samples, the accuracy of the bootstrap methods for constructing one-sided confidence bounds are quite different for different quantiles when Type I censored data are considered. The problem occurs when the quantity of interest is the p quantile where p is close to the proportion failing. The reason for the problem is due to the discrete-like behavior of the ML estimator of such quantiles in Type I censored data (see Jeng and Meeker (2000) for more discussion and examples on this point). When the expected number of failures is small (less than 10), another alternative suggested by some limited simulation results is to use a double bootstrap calibration. Both the theoretical and the finite sample properties of this approach could be studied. The computational effort needed to do a complete simulation experiment would, however, be extremely large.

7 Proofs

Let $C, C_1, C_2, \ldots$ denote generic positive constants that do not depend on n. Unless otherwise mentioned, limits in the order symbols $O(\cdot)$ and $o(\cdot)$ are taken by letting $n \to \infty$.

Proof of Proposition 1: Define the measures $\mu_n, n \in \mathbb{Z}_+$ on the Borel σ-field $\mathcal{B}(\mathbb{R}^m)$ on $\mathbb{R}^m$ by

$$\mu_n(A) = \int_{A \cap B_0} f_n d\lambda, \quad A \in \mathcal{B}(\mathbb{R}^m). \tag{7.1}$$

Then, $\mu_n, n \geq 0$ are finite measures and by Condition (ii), $\mu_n(\mathbb{R}^m) \to \mu_0(\mathbb{R}^m)$. Hence, by an extended version of Scheffe's Theorem (cf. p.215, Billingsley (1995)),

$$\begin{aligned} &\sup\left\{\int_{B_0} e^{i(tx+vy)} f_n(x,y) d\lambda(x,y) - \int_{B_0} e^{i(tx+vy)} f_0(x,y) d\lambda(x,y) : (t,v) \in \mathbb{R}^m\right\} \\ &\leq \int_{B_0} |f_n - f| d\lambda \to 0 \text{ as } n \to \infty. \end{aligned} \tag{7.2}$$

Fix $\epsilon > 0$ and write $x_0 = (x_{01}, \ldots, x_{0m_1})$. Then, uniformly in $(t,v) \in \mathcal{C}(\epsilon)$,

$$\begin{aligned} & \left|\int_{B_0} e^{i(tx+vy)} f_0(x,y) d\lambda(x,y)\right| \\ \leq \;& c\left|\sum_{y \in [l_1, l_2]} m_2 \int_{\mathcal{O}} e^{itx} dx \cdot e^{ivy}\right| + \int_{B_0} [f_0(x,y) - c] d\lambda(x,y) \\ = \;& c\left|\prod_{j=1}^{m_1} \int_{\mathcal{O}} e^{it(x-x_0)} dx\right| \cdot \left|\prod_{j=1}^{m_2} \sum_{k=l_1}^{l_2} e^{iv_{jk}}\right| + \int_{B_0} f_0(x,y) d\lambda(x,y) - c\lambda(B_0) \\ = \;& c\lambda(B_0)\left\{\prod_{j=1}^{m_1}\left|\int_{-a_j}^{a_j} \frac{e^{it_j u}}{2a_j} du\right|\right\}\left\{\prod_{j=1}^{m_2}\left|(l_2 - l_1 + 1)^{-1} \sum_{k=l_1}^{l_2} e^{iv_{jk}}\right|\right\} + \mu_0(\mathbb{R}^m) - c\lambda(B_0) \\ < \;& \mu_0(\mathbb{R}^m) - \epsilon_1 \end{aligned} \tag{7.3}$$

for some constant $\epsilon_1 > 0$, depending on $\epsilon, c, \lambda(B_0)$. This follows from properties of characteristic functions and from the facts that in the second line above, the terms under the first product are characteristic functions of the UNIFORM$(-a_j, a_j)$ distributions, $j = 1, \ldots, m$ and those under the second product are the characteristic function of the discrete uniform distribution over the integers $[l_1, l_1 + 1, \ldots, l_2]$, having maximal span 1 (cf. Ch. 15, Feller (1968)).

Next, define the measures $\gamma_n, n \geq 0$ by

$$P((X_n, Y_n) \in A) = \mu_n(A) + \gamma_n(A), A \in \mathcal{B}(\mathbb{R}^m). \tag{7.4}$$

Because $f_0(x,y) > c$ on B_0, it follows that $\mu_0(\mathbb{R}^m) = \mu_0(B_0) \in (0,1]$, and hence $\gamma_0(\mathbb{R}^m) < 1$. Let $\gamma = \epsilon_1/3$. By (7.2) there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$\int_{B_0} |f_n - f| d\lambda < \delta \text{ and } \gamma_n(\mathbb{R}^m) < \gamma_0(\mathbb{R}^m) + \delta.$$

Then, for all $n \geq n_0$,

$$
\begin{aligned}
& \sup \left\{ |Ee^{itX_n + ivY_n}| : (t,v) \in \mathcal{C}(\epsilon) \right\} \\
\leq \;& \gamma_n(I\!R^m) + \sup \left\{ \left| \int e^{itx+ivy} d\mu_n(x,y) \right| : (t,v) \in \mathcal{C}(\epsilon) \right\} \\
\leq \;& \left[\gamma_0(I\!R^m) + \frac{\delta}{4} \right] + \int_{B_0} |f_n - f_0| d\lambda \\
& + \sup \left\{ \left| \int e^{itx+ivy} d\mu_0(x,y) \right| : (t,v) \in \mathcal{C}(\epsilon) \right\} \\
\leq \;& \gamma_0(I\!R^m) + \delta + \delta + \mu_0(I\!R^m) - \epsilon_1 < 1 - \delta.
\end{aligned}
$$

Hence, Proposition 1 follows.

Lemma 1: *Let $(X_{ni}, Y_{ni}) \in I\!R^{p+q}, i = 1, \ldots, n, n \geq 1$ be a triangular array of row iid random vectors such that for each $n \geq 1, Y_{n1} \in I\!R^q$ is a lattice variable with minimal lattice $Z\!\!\!Z^q, EX_{n1} = 0, EY_{n1} = 0$ and $Cov(X_{n1}, Y_{n1}) = I_{p+q}$, the identity matrix of order $p+q$. Suppose that*

(i) there exists a $\delta \in (0,1)$ and an integer $s \geq 3$ such that for all $n > \delta^{-1}$,

$$E \parallel X_{n1} \parallel^{\alpha(s)} \leq \delta^{-1}, \qquad E \parallel Y_{n1} \parallel^{\beta(s)} \leq \delta^{-1}, \tag{7.5}$$

for $\alpha(s) = \max\{2s+1, p+1\}$ and $\beta(s) = \max\{\alpha(s), q+1\}$;

(ii) for any $\epsilon > 0$, there exists a $\delta \in (0,1)$ such that for all $n \geq \delta^{-1}$,

$$\sup\{|E \exp(i(tX_{n1} + vY_{n1}))| : (t,v) \in \mathcal{C}(\epsilon)\} \leq 1 - \delta. \tag{7.6}$$

Let $g : I\!R^{p+q} \to I\!R$ be $(s-1)$-times continuously differentiable in a neighborhood of $0 \in I\!R^{p+q}$ with $g(0) = 0, \partial_j g(0) \neq 0$ for some $1 \leq j \leq p$ and $\sum_{j=1}^p [\partial_j g(0)]^2 = 1$.
Then, there exist functions $p_j(\cdot;\cdot)$ such that

$$
\begin{aligned}
& \sup_{u \in I\!R} \left| P\left(\sqrt{n} g\left(\frac{S_n}{\sqrt{n}} \right) \leq u \right) - \int_{-\infty}^u \left[1 + \sum_{j=1}^{s-2} n^{-j/2} p_j(u; \{\chi_{\nu n} : |\nu| \leq j+2\}) \right] \phi(u) du \right| \\
=\;& O(n^{-(s-1)/2}) \qquad (7.7)
\end{aligned}
$$

where $S_n = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ni}, \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_{ni} \right)$, where $\chi_{\nu n}$ is the νth cumulant of $(X_{n1}, Y_{n1}), \nu \in Z\!\!\!Z_+^{p+q}$.
Proof: A version of this result for sequences (as opposed to triangular arrays) may be deduced by using Theorem 1 of Jensen (1989). Here we outline the extensions of his arguments needed to handle the triangular array case. Without loss of generality suppose that Conditions (i) and (ii) of Lemma 1 hold for all $n \geq 1$. The first step in the proof involves deriving an expansion for the conditional probability distribution of $S_n^{(1)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ni}$, given $S_n^{(2)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_{ni}$.

By Bartlett (1938)'s formula for conditional characteristic function,

$$E[\exp(itS_n^{(1)})|S_n^{(2)}=y] = (2\pi)^{-q}\int_{[-\pi\sqrt{n},\pi\sqrt{n}]^q}\left[f_n^n\left(\frac{t}{\sqrt{n}},\frac{v}{\sqrt{n}}\right)\right]e^{-ivy}dv/a_n \tag{7.8}$$

with $a_n = n^{q/2}P(S_n^{(2)}=y)$ and $f_n(t,v)\equiv E\exp(itX_{n1}+ivY_{n1}), t\in I\!R^p, v\in I\!R^q$. Let $\psi_n(t,v)$ be the Fourier transform of the Edgeworth expansion for $(S_n^{(1)},S_n^{(2)})$, given by

$$\psi_n(t,v) = e^{-(\|t\|^2+\|v\|^2)/2}\left[1+\sum_{j=1}^{\alpha(s)-3}n^{-j/2}\tilde{P}_j(it,iv;\{\chi_{\nu n}:|\nu|\le j+2\})\right] \tag{7.9}$$

where the function $\tilde{P}_j(it,iv;\{\chi_{\nu n}:|\nu|\le j+2\}), j\ge 1$ are defined in the usual way (namely, by identity (7.2) of [BR] with χ_ν's there replaced by $\chi_{\nu n}$'s).

Let $\hat{\Psi}_n(t|y)$ be defined, in analogy to (7.8), by

$$\hat{\Psi}_n(t|y) = (2\pi)^{-q}\int \psi_n(t,v)e^{-ivy}dv/b_n \tag{7.10}$$

where $b_n \equiv (2\pi)^{-q}\int\psi_n(0,v)e^{-ivy}du$. By Condition (i) and arguments in (2.9)-(2.12) of Jensen (1989), it follows that uniformly in $\| y \|^2\le s\log n$,

$$|a_n-b_n| = O(n^{-(\alpha(s)-2)/2})$$

and $a_n\bigwedge b_n\ge C_1n^{-s/2}$ for some $C_1\in(0,\infty)$, so that

$$|a_n-b_n|a_n^{-1} = O(n^{-(\alpha(s)-2-s)/2}) = O(n^{-(s-1)/2}). \tag{7.11}$$

Arguments leading to Theorem 9.9 of [BR] yield

$$\left|\partial^\nu(f_n^n(\frac{t}{\sqrt{n}},\frac{v}{\sqrt{n}})-\psi_n(t,v))\right| \le C_2(\delta)n^{-(\alpha(s)-2)/2}\left[1+\| t \|^{3(\alpha(s)-2)+|\nu|}\right]e^{-(\|t\|^2+\|u\|^2)/4} \tag{7.12}$$

for $|\nu|\le\alpha(s)$, and for $\| t \| + \| v \|\le C_3(\delta)\sqrt{n}$. Now using the smoothing inequality of Corollary 11.2 of [BR], and arguments in the proof of Lemma 1 of Jensen (1989), one gets

$$\sup_{B\in\mathcal{B}}\left|P(S_n^{(1)}\in B|S_n^{(2)}=y)-\Psi_n(B|y)\right| = O(n^{-(s-1)/2}) \tag{7.13}$$

uniformly over $\sqrt{n}y\in Z\!\!\!Z^{m_2}$ with $\| y \|\le s\log n$, where $\Psi(\cdot|y)$ is the signed measure corresponding to the Fourier transform $\hat{\Psi}(\cdot|y)$ of (7.10) and $\mathcal{B}$ is any given collection of Borel subsets of $I\!R^p$ satisfying

$$\Phi([\partial B]^\epsilon) = O(\epsilon) \tag{7.14}$$

as $\epsilon\downarrow 0$.

Next, using the transformation technique of Bhattacharya and Ghosh (1978), (hereafter referred to as [BG]) one can easily show that uniformly over $\| y \|^2\le s\log n$,

$$\begin{aligned}&\sup_{u_0\in I\!R}\left|P\left(\sqrt{n}g\left(\frac{S_n}{\sqrt{n}}\right)\le u_0|S_n^{(2)}=y\right)-\int_{-\infty}^{u_0}\left[1+\sum_{j=1}^{s-2}n^{-1/2}\check{p}_{jn}(u;y)\right]\phi_{\sigma_n}(u-d_ny)du\right| \\ =\ & O(n^{-(s-1)/2}) \end{aligned}\tag{7.15}$$

for some polynomials $\check{p}_{jn}(\cdot; y)$, whose coefficients are rational functions of $\{\chi_{\nu n} : |\nu| \leq j+2\}$ and $\{\partial^\nu g(0; y/\sqrt{n}) : |\nu| \leq s-1\}$. Here $\sigma_n^2 = \sum_{j=1}^p [\partial_j g(0, y/\sqrt{n})]^2$ and d_n is the $1 \times q$ vector with ith component $\partial_{p+i} g(0, y/\sqrt{n}), i = 1, \ldots, q$. Next, note that $P(\sqrt{n} g(S_n/\sqrt{n}) \leq u_0) = E\{P(\sqrt{n} g(S_n/\sqrt{n}) \leq u_0 | S_n^{(2)})\} = E[\int_{-\infty}^{u_0} \{1 + \sum_{j=1}^{s-2} n^{-j/2} \check{p}_{jn}(u; S_n^{(2)})\} \phi_{\sigma_n}(u - d_n S_n^{(2)}) du] + O(P(\| S_n^{(2)} \|^2 > s \log n) + n^{-(s-1)/2})$. Hence, using the arguments in Göetz and Hipp (1978) for expansions of expectations of smooth functions and for probabilities of moderate deviations, and using an analog of (7.12) with $t = 0, |\nu| \leq \beta(s)$ under the moment condition (i) on $Y_{ni}'s$, one gets (7.7).

Proof of Theorem 1: By using a Taylor's expansion of the left side of equation (2.3) around θ_n up to order s, one can express (2.3) as

$$0 = \partial_j \bar{l}_n(\theta) = \partial_j \bar{l}_n(\theta_n) + \sum_{|\nu|=1}^{s-1} [\partial^\nu \partial_j \bar{l}_n(\theta_n)](\theta - \theta_n)^\nu / \nu! + R_{nj}, \tag{7.16}$$

where $|R_{nj}(\theta)| \leq C|\theta - \theta_n|^s \cdot \sup\{|\partial^\nu \bar{l}_n(t)| : \| t - \theta_n \| \leq \| \theta - \theta_n \|, |\nu| = s+1\},\ 1 \leq j \leq k$.

Now, using Assumption (A.2) and (A.3) and Corollary 4.2 of Fuk and Nagaev (1971), we have

$$\begin{aligned} &P_n(|\partial^\nu \bar{l}_n(\theta_n) - E_n \partial^\nu \bar{l}_n(\theta_n)| > C n^{-1/2} (\log n)^{1/2}) \\ &= O(n^{-(s-2)/2} (\log n)^{-s/2}), 1 \leq |\nu| \leq s-1 \end{aligned} \tag{7.17}$$

and

$$P_n\left(\sup\left\{|\partial^\nu \bar{l}_n(\theta)| : \| \theta - \theta_n \| \leq a_1, |\nu| = s+1\right\} > C\right) = O(n^{-(s-2)/2} (\log n)^{-s/2}). \tag{7.18}$$

Hence, on a set A_n with $P_n(A_n^c) = O(n^{-(s-2)/2} (\log n)^{-s/2})$, we may rewrite (7.16) as

$$(\theta - \theta_n) = g_n(\theta - \theta_n) \tag{7.19}$$

for some continuous function g_n that satisfies $\| g_n(x) \| \leq C n^{-1/2} (\log n)^{1/2}$ for all $\| x \| \leq C n^{-1/2} (\log n)^{1/2}$. Hence, part (a) follows from Brouwer's Fixed Point theorem, as in the proof of Theorem 3 of [BG].

To prove part (b), note that using the arguments in the proof of Theorem 3 of [BG], we can express $\hat{\theta}_n$ and θ_n as

$$\hat{\theta}_n = g(\bar{Z}_n^\dagger) \text{ and } \theta_n = g(E_n \bar{Z}_n) \tag{7.20}$$

for some smooth function $g : I\!R^m \to I\!R^k$ where $Z_{in}^{\dagger(\nu)} = Z_{in}^{(\nu)}$ for $|\nu| \geq 2$ and $Z_{in}^{\dagger(\nu)} = Z_{in}^{(\nu)} + R_n(\hat{\theta}_n), |\nu| = 1$, and $R_n(\hat{\theta}_n)$ is the vector of $R_{nj}(\hat{\theta}_n), 1 \leq j \leq k$. Now using the reparametrization of the Z_{in}'s in terms of $\tilde{Z}_{in}$'s and using Lemma 1 above in place of Theorem 1 of Jensen (1989), one can complete the proof of part (b) as in the proof of Theorem 2.1 of Jensen (1993). We omit the routine details.

Proof of Theorem 2: Following the arguments on page 8-9 of Jensen (1993), we can express $\tilde{R}_n$ and $\tilde{W}_n$ as $\tilde{R}_n = \tilde{V}_{1n}$ and $\tilde{W}_n = \tilde{V}_{2n}^2$ where $\tilde{V}_{1n}$ and $\tilde{V}_{2n}$ admit stochastic expansions of the form, for $m = 1, 2$,

$$\tilde{V}_{mn} \quad = \quad \sum_{i=k_1+1}^{k} a_{mi} \left[\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tilde{Z}_{in}^{(e_i)}\right] + \sum_{r=1}^{2} n^{-r/2} \check{p}_{rm} \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \tilde{Z}_{in}; \theta_n)\right) + \check{R}_{mn} \tag{7.21}$$

for some constants $a_{mi} = a_{min} \in I\!R| \setminus \{0\}$ and polynomials $\check{p}_{rm}(\cdot;\cdot)$ (with $k_1 = k-1$ when $m = 1$), such that the remainder terms satisfy the inequality

$$P_n(|\check{R}_{mn}| > Cn^{-1}(\log n)^{-2}) = O(n^{-1}(\log n)^{-2}). \tag{7.22}$$

Now applying Lemma 1 above, and the transformation techniques of [BG], one can establish parts (a) and (b) of Theorem 2. The proof for $\tilde{W}_n$ is similar, by noting that the effect of the correction factor $\frac{1}{1+B(\cdot)}$ shows up only in the term of order $O(n^{-1})$ in the expansion for $\tilde{W}_n$.

Proof of Theorem 3. By part (a) of Theorem 1, under (A.4)(i) with $s = 3$, $E_0 \parallel Z_{10} \parallel^{2s+1} \leq \parallel A_0 \parallel^{2s+1} E_0 \parallel \tilde{Z}_{10} \parallel^{2s+1} < \infty$, so that

$$P_0(\parallel \hat{\theta}_n - \theta_0 \parallel > a_1[(\log n)/n]^{1/2}) = o(n^{-5/2}). \tag{7.23}$$

Hence, by the Borel-Cantelli lemma, $\hat{\theta}_n - \theta_0 = O(n^{-1/2}(\log n)^{1/2})$ a.s. (P_0). Let D_0 be the set of P_0-probability 1 where $\hat{\theta}_n - \theta_0 = O(n^{-1/2}(\log n)^{1/2})$ as $n \to \infty$. Then, by the continuity of $\partial^\nu l(x;\theta)$ in θ over Θ, $1 \leq |\nu| \leq s$, and the continuity of the second moments, $f(x, y; \hat{\theta}_n) \to f(x, y; \theta_0)$ as $n \to \infty$ for all $(x, y) \in I\!R^{m_1} \times Z\!\!\!Z^{m_2}$. Hence, the conditions of Proposition 1 hold, which in turn implies (A.4)(ii) along every realization of $\{\hat{\theta}_n\}$ on D_0. Now, using the expansion for $\sqrt{n}(\hat{\theta}_n - \theta_0)$ from part (b) of Theorem 1 and the transformation technique of [BG], one can show that

$$\sup_{B\in\mathcal{B}} \left| P(T_n^* \in B) - \int_B \left[1 + q_1(x;\hat{\theta}_n)n^{-1/2}\right] \phi(x)dx \right| = o(n^{-1/2}) \quad a.s.(P_0) \tag{7.24}$$

and

$$\sup_{B\in\mathcal{B}} \left| P(T_n \in B) - \int_B \left[1 + q_1(x,\theta_0)n^{-1/2}\right] \phi(x)dx \right| = o(n^{-1/2}) \tag{7.25}$$

for some polynomial $q_1(\cdot) \equiv q_1(\cdot;\theta)$ with coefficients that are smooth functions of θ. Part (a) of the theorem follows from this. Part (b) follows by similar arguments, by exploiting the continuity of the cumulants of $\tilde{Z}_1(\theta)$ in θ and the reparametrization argument in Remark 2.5 of Jensen (1993). We omit the details.

Appendix

This appendix demonstrates that the (A.1)$'$-(A.3)$'$ conditions hold for Equation (5.8). It is clear that if Φ_{LS} has a ν-th derivative w.r.t. θ on $I\!R$ x Θ, then Condition (A.1)$'$ holds. This is true for the SEV, normal, and logistic distributions.

For condition (A.2)$'$, we present the general formulation for location-scale distributions and then discuss the details for the SEV, normal, and logistic distributions.

Let $\xi_i = (x_i - x_p)/\sigma + u_p$, then

$$\frac{\partial \xi_i}{\partial x_p} = -\frac{1}{\sigma}, \quad \frac{\partial^j \xi_i}{\partial x_p^j} = 0, \quad j \geq 2, \qquad \frac{\partial \xi_i}{\partial \sigma} = -\frac{x_i - x_p}{\sigma^2}, \quad \frac{\partial^j \xi_i}{\partial \sigma^j} = (-1)^j \frac{x_i - x_p}{\sigma^{j+1}}, \quad j \geq 2,$$
$$\frac{\partial^{j+k} \xi_i}{\partial x_p^j \partial \sigma^k} = 0, \quad j \geq 2, k \geq 1, \quad \frac{\partial^{j+k} \xi_i}{\partial x_p^j \partial \sigma^k} = \left(-\frac{1}{\sigma}\right)^{k+1}, \quad j = 1, k \geq 1. \tag{H.1}$$

Let $\xi_c = (x_c - x_p)/\sigma + u_p$ denote the standardized censoring time, then

$$\frac{\partial \xi_c}{\partial x_p} = -\frac{1}{\sigma}, \quad \frac{\partial^j \xi_c}{\partial x_p^j} = 0, \quad j \geq 2, \qquad \frac{\partial \xi_c}{\partial \sigma} = -\frac{x_c - x_p}{\sigma^2}, \quad \frac{\partial^j \xi_c}{\partial \sigma^j} = (-1)^j \frac{x_c - x_p}{\sigma^{j+1}}, \quad j \geq 2,$$
$$\frac{\partial^{j+k} \xi_c}{\partial x_p^j \partial \sigma^k} = 0, \quad j \geq 2, k \geq 1, \frac{\partial^{j+k} \xi_c}{\partial x_p^j \partial \sigma^k} = \left(-\frac{1}{\sigma}\right)^{k+1}, \quad j = 1, k \geq 1. \tag{H.2}$$

The partial derivatives of the log likelihood function are

$$\begin{aligned}
\partial^{(0,1)} l(x_i; (\sigma, x_p)) &= \delta_i \left[\frac{\phi'_{\mathrm{LS}}(\xi_i) \frac{\partial \xi_i}{\partial x_p}}{\phi_{\mathrm{LS}}(\xi_i)} \right] + (1 - \delta_i) \left[- \frac{\phi_{\mathrm{LS}}(\xi_c) \frac{\partial \xi_c}{\partial x_p}}{1 - \Phi_{\mathrm{LS}}(\xi_c)} \right], \\
\partial^{(1,0)} l(x_i; (\sigma, x_p)) &= \delta_i \left[-\frac{1}{\sigma} + \frac{\phi'_{\mathrm{LS}}(\xi_i) \frac{\partial \xi_i}{\partial \sigma}}{\phi_{\mathrm{LS}}(\xi_i)} \right] + (1 - \delta_i) \left[- \frac{\phi_{\mathrm{LS}}(\xi_c) \frac{\partial \xi_c}{\partial \sigma}}{1 - \Phi_{\mathrm{LS}}(\xi_c)} \right], \\
\partial^{(1,1)} l(x_i; (\sigma, x_p)) &= \delta_i \left\{ \frac{\phi_{\mathrm{LS}}(\xi_i) [\phi''_{\mathrm{LS}}(\xi_i) \frac{\partial \xi_i}{\partial x_p} \frac{\partial \xi_i}{\partial \sigma} + \phi'_{\mathrm{LS}}(\xi_i) \frac{\partial^2 \xi_i}{\partial x_p \partial \sigma}] - [\phi'_{\mathrm{LS}}(\xi_i)]^2 \frac{\partial \xi_i}{\partial x_p} \frac{\partial \xi_i}{\partial \sigma}}{\phi_{\mathrm{LS}}(\xi_i)^2} \right\} \\
&+ (1 - \delta_i) \left\{ \frac{[1 - \Phi_{\mathrm{LS}}(\xi_c)][-\phi'_{\mathrm{LS}}(\xi_c) \frac{\partial \xi_c}{\partial x_p} \frac{\partial \xi_c}{\partial \sigma} + \phi_{\mathrm{LS}}(\xi_c) \frac{\partial^2 \xi_c}{\partial x_p \partial \sigma}] - [\phi_{\mathrm{LS}}(\xi_c)]^2 \frac{\partial \xi_c}{\partial x_p} \frac{\partial \xi_c}{\partial \sigma}}{[1 - \Phi_{\mathrm{LS}}(\xi_c)]^2} \right\}, \\
\partial^{(0,2)} l(x_i; (\sigma, x_p)) &= \delta_i \left\{ \frac{\phi_{\mathrm{LS}}(\xi_i) [\phi''_{\mathrm{LS}}(\xi_i) (\frac{\partial \xi_i}{\partial x_p})^2] - [\phi'_{\mathrm{LS}}(\xi_i)]^2 (\frac{\partial \xi_i}{\partial x_p})^2}{\phi_{\mathrm{LS}}(\xi_i)^2} \right\} \\
&+ (1 - \delta_i) \left\{ \frac{[1 - \Phi_{\mathrm{LS}}(\xi_c)][\phi'_{\mathrm{LS}}(\xi_c) (\frac{\partial \xi_c}{\partial x_p})^2] - [\phi_{\mathrm{LS}}(\xi_c)]^2 (\frac{\partial \xi_c}{\partial x_p})^2}{[1 - \Phi_{\mathrm{LS}}(\xi_c)]^2} \right\}, \\
\partial^{(2,0)} l(x_i; (\sigma, x_p)) &= \delta_i \left\{ \frac{\phi_{\mathrm{LS}}(\xi_i) [\phi''_{\mathrm{LS}}(\xi_i) (\frac{\partial \xi_i}{\partial \sigma})^2] - [\phi'_{\mathrm{LS}}(\xi_i)]^2 (\frac{\partial \xi_i}{\partial \sigma})^2}{\phi_{\mathrm{LS}}(\xi_i)^2} \right\} \\
&+ (1 - \delta_i) \left\{ \frac{[1 - \Phi_{\mathrm{LS}}(\xi_c)][\phi'_{\mathrm{LS}}(\xi_c) (\frac{\partial \xi_c}{\partial \sigma})^2] - [\phi_{\mathrm{LS}}(\xi_c)]^2 (\frac{\partial \xi_c}{\partial \sigma})^2}{[1 - \Phi_{\mathrm{LS}}(\xi_c)]^2} \right\}.
\end{aligned} \tag{H.3}$$

Similarly, for $3 \leq |\nu| \leq 4$, $\partial^\nu l(x_i; (\sigma, x_p))$ is a function of $\Phi_{\mathrm{LS}}, \phi_{\mathrm{LS}}, \phi'_{\mathrm{LS}}, \phi''_{\mathrm{LS}}, \phi'''_{\mathrm{LS}}, \phi''''_{\mathrm{LS}}$, and terms in the Equation (H.1), (H.2). This establishes Condition (A.2)$'$.

SEV Distribution. For the smallest extreme value distribution,

$$\Phi_{\mathrm{LS}}(\xi) = 1 - \exp\{[-\exp(\xi)]\}, \quad \phi_{\mathrm{LS}}(\xi) = \exp\{[\xi - \exp(\xi)]\}, \quad \phi'_{\mathrm{LS}}(\xi) = [1 - \exp(\xi)]\phi(\xi).$$

Then Equations (H.3) becomes:

$$\begin{aligned}
\partial^{(0,1)} l_i &= \delta_i \left[(1 - \exp(\xi_i)) \frac{\partial \xi_i}{\partial x_p} \right] + (1 - \delta_i) \left[- \exp(\xi_c) \frac{\partial \xi_c}{\partial x_p} \right], \\
\partial^{(1,0)} l_i &= \delta_i \left[-\frac{1}{\sigma} + (1 - \exp(\xi_i)) \frac{\partial \xi_i}{\partial \sigma} \right] + (1 - \delta_i) \left[- \exp(\xi_c) \frac{\partial \xi_c}{\partial \sigma} \right].
\end{aligned} \tag{H.4}$$

From Equations (H.1), (H.2), and (H.4), $\partial^{(0,1)}l$ and $\partial^{(1,0)}l$ are linearly independent. We can see that all other partial derivatives $\partial^{\nu}l$, $1 \le |\nu|$, are functions of $\exp(\xi_i)$ and the terms in Equations (H.1), (H.2). Then $\tilde{Z}_i$ can be written as

$$\left(\partial^{(0,1)}l_i,\, \partial^{(1,0)}l_i,\, \delta_i\xi_i\exp(\xi_i),\, \delta_i\xi_i^2\exp(\xi_i),\, \delta_i\xi_i^3\exp(\xi_i),\, \delta_i\xi_i^4\exp(\xi_i),\, \delta_i\right).$$

Because the expectations of $\xi_i^j\exp(\xi_i)^k$, $0 < j+k$, $0 \le k \le 1$, are finite over an open set containing the true parameters, the expectations of $\partial^{\nu}l$, $1 \le |\nu|$, are finite over the same open set. This establishes Condition (A.2)$'$.

Normal Distribution. For the normal distribution

$$\phi_{\mathrm{LS}}(\xi) = \frac{1}{\sqrt{\pi}}e^{-\xi^2/2}, \quad \Phi_{\mathrm{LS}}(\xi) = \int_{-\infty}^{\xi}\phi(x)dx, \quad \phi_{\mathrm{LS}}'(\xi) = -\xi\phi(\xi).$$

Then Equations (H.3) becomes:

$$\begin{aligned}
\partial^{(0,1)}l_i &= \delta_i\left[\xi_i\frac{\partial\xi_i}{\partial x_p}\right] + (1-\delta_i)\left[-\frac{\phi(\xi_c)\frac{\partial\xi_c}{\partial x_p}}{1-\Phi(\xi_c)}\right], \\
\partial^{(1,0)}l_i &= \delta_i\left[-\frac{1}{\sigma} + \xi_i\frac{\partial\xi_i}{\partial\sigma}\right] + (1-\delta_i)\left[-\frac{\phi(\xi_c)\frac{\partial\xi_c}{\partial\sigma}}{1-\Phi(\xi_c)}\right].
\end{aligned} \tag{H.5}$$

From Equations (H.1), (H.2) and (H.5), $\partial^{(0,1)}l$, and $\partial^{(1,0)}l$ are linearly independent. We can see that all of the other partial derivatives $\partial^{\nu}l$, $1 \le |\nu|$, are functions of ξ_i, and terms in (H.1) and (H.2). Then $\tilde{Z}_i$ can be written as

$$\left(\partial^{(0,1)}l_i,\, \partial^{(1,0)}l_i,\, \delta_i\xi_i^3,\, \delta_i\xi_i^4,\, \delta_i\right).$$

Because the expectations of ξ_i^j, $0 < j$, are finite over an open set containing the true parameters, the expectations of $\partial^{\nu}l$, $1 \le |\nu|$, are finite over the same open set. This establishes Condition (A.2)$'$.

Logistic Distribution. For the logistic distribution

$$\Phi_{\mathrm{LS}}(\xi) = \frac{1}{1+e^{-\xi}}, \quad \phi_{\mathrm{LS}}(\xi) = \frac{e^{-\xi}}{(1+e^{-\xi})^2}, \quad \phi_{\mathrm{LS}}'(\xi) = -\Phi_{\mathrm{LS}}(\xi)\phi_{\mathrm{LS}}(\xi).$$

Then from Equation (H.3):

$$\begin{aligned}
\partial^{(0,1)}l_i &= \delta_i\left[-\Phi_{\mathrm{LS}}(\xi_i)\frac{\partial\xi_i}{\partial x_p}\right] + (1-\delta_i)\left[-\Phi_{\mathrm{LS}}(\xi_c)\frac{\partial\xi_c}{\partial x_p}\right], \\
\partial^{(1,0)}l_i &= \delta_i\left[-\frac{1}{\sigma} - \Phi_{\mathrm{LS}}(\xi_i)\frac{\partial\xi_i}{\partial\sigma}\right] + (1-\delta_i)\left[-\Phi_{\mathrm{LS}}(\xi_c)\frac{\partial\xi_c}{\partial\sigma}\right].
\end{aligned} \tag{H.6}$$

From Equation (H.1), (H.2) and (H.6), $\partial^{(0,1)}l$, and $\partial^{(1,0)}l$ are linearly independent. We can see that all other partial derivatives $\partial^{\nu}l$, $1 \le |\nu|$, are functions of $\Phi_{\mathrm{LS}}(\xi_i)$, $\phi_{\mathrm{LS}}(\xi_i)$, and terms in the Equations (H.1) and (H.2). Then $\tilde{Z}_i$ can be written as

$$\Big(\partial^{(0,1)}l_i,\ \partial^{(1,0)}l_i,\ \delta_i\phi_{\mathrm{LS}}(\xi_i),\ \delta_i\phi_{\mathrm{LS}}^{'}(\xi_i),\ \delta_i\phi_{\mathrm{LS}}^{''}(\xi_i),\ \delta_i\phi_{\mathrm{LS}}(\xi_i)\xi_i,\ \delta_i\phi_{\mathrm{LS}}^{'}(\xi_i)\xi_i,\ \delta_i\phi_{\mathrm{LS}}^{''}(\xi_i)\xi_i,$$
$$\delta_i\phi_{\mathrm{LS}}^{'}(\xi_i)\xi_i^2,\ \delta_i\phi_{\mathrm{LS}}^{''}(\xi_i)\xi_i^2,\ \delta_i\phi_{\mathrm{LS}}^{''}(\xi_i)\xi_i^3,\ \delta_i\phi_{\mathrm{LS}}^{''}(\xi_i)\xi_i^4,\ \delta_i\Big) \tag{H.7}$$

Because the expectations of $\xi_i^j\phi_{\mathrm{LS}}(\xi_i)^k\Phi_{\mathrm{LS}}(\xi_i)^m$, $j+k+m>0$, are finite over an open set containing the true parameters, the expectations of $\partial^\nu l$, $1 \leq |\nu|$, are finite over the same open set. Thus Condition (A.2)$'$ holds.

For right censoring and a location-scale distribution with a likelihood function satisfying Conditions (A.1)$'$ and (A.2)$'$, it can be shown by using Equation (H.3) that $I(\theta_0) = D(\theta_0)$, $\theta_0 \in \Theta_0 \subset \Theta$. The calculation is straight forward, we omit the detail here. Note that D is the variance-covariance matrix of score function so it is nonnegative definite. If the determinant of D is 0, then

$$\frac{\partial l(X_i;(\sigma,x_p))}{\partial x_p} = c\frac{\partial l(X_i;(\sigma,x_p))}{\partial\sigma}, \tag{H.8}$$

for all possible values of X_i, where c is a constant. From Equations (H.4), (H.5), and (H.6) we see that (H.8) is not true for the SEV, normal, and logistic distributions. Thus D is positive definite. Thus Condition (A.3)$'$ holds.

References

Babu, G. J. (1991), Edgeworth expansions for statistics which are functions of lattice and non-lattice variables, *Statistics & Probability Letters*, **12**, 1-7.

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.

Bartlett, M. S. (1938), The characteristic function of a conditional statistic, *Journal of London Math Society* **13**, 62-67.

Bhattacharya, R.N. and Ghosh, J.K. (1978). Validity of the formal Edgeworth expansion, *Annals of Statistics*, **6**, 434-451.

Bhattacharya, R.N. and Ranga Rao, R. (1986), *Normal Approximation and Asymptotic Expansions*, Krieger Publishing Company, Inc. (Melbourne, FL).

Billingsley, P. (1995). *Probability and Measure*, 3rd Edition, Wiley, New York, NY.

Box, G. E. P. (1949), A general distribution theory for a class of likelihood criteria, *Biometrika*, **36**, 317-346.

Chandra, T. K. and Ghosh, J. K. (1979), Valid asymptotic expansions for the likelihood ratio statistics and other perturbed chi-square variables, *Sankhyā A*, **41**, 22-47.

Datta, S. (1992), A note on the Edgeworth expansions and the bootstrap, *Sankhyā A*, **54**, 171-182.

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

Efron, B. (1985), Bootstrap confidence intervals for a class of parameter problems, *Biometrika*, **72**, 45-58.

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

Feller, W. (1968), *An Introduction to Probability Theory and its Applications*, Volume 2, 3rd Edition, Wiley, New York, NY.

Fuk, D. Kh. , and Nagaev, S. V. (1971), Probability inequalities for sums of independent random variables, *Theory of Probability and its Applications*, **16**, 643-660.

Götze, F., and Hipp, C. (1978), Asymptotic expansions in the central limit theorem under moment conditions, , *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **42**, 67-87.

Gross, S.T. and Lai, T.L. (1996), Bootstrap methods for truncated and censored data, *Statistica Sinica* **6**, 509-530.

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

Horvath, L. and Yandell, B.S. (1987), Convergence rates for bootstrapped product-limit process, *Annals of Statistics*, **15**, 1155-1173.

Jeng, S. L. and Meeker, W. Q. (2000), Comparisons of approximate confidence interval procedures for type I censored data", *Technometrics*, **42**,135-148.

Jensen, J. L. (1987), Standardized log-likelihood ratio statistics for mixtures of discrete and continuous observations, *The Annals of Statistics*, **15**, 314-324.

Jensen, J. L. (1989), Validity of the formal Edgeworth expansion when the underlying distribution is partly discrete, *Probability Theory and Related Fields*, **81**, 507-519.

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.

Lawley, D. N. (1956), A general method for approximating to the distribution of the likelihood ratio criteria, *Biometrika*, **43**, 295-303.

Lai, T.L. and Wang, J.Q. (1993), Edgeworth expansions for symmetric statistics with applications to bootstrap methods, *Statistica Sinica*, **3**, 517-542.

Lehmann, E.L. (1986), *Theory of Point Estimation*, Wiley, New York, NY.

Lo, S.H. and Singh, K. (1986), The product-limit estimator and the bootstrap: Some asymptotic representations, *Probability Theory and Related Fields*, **71**, 455-465.

McCullagh, P. (1984) Local sufficiency, *Biometrika*, **71**, 233-244.

Nishii, R. and Yanagimoto, T. (1993), Normal approximation to the distribution of the sample mean in the exponential family , *Statistical Sciences and Data Analysis*, (Tokyo, 1991), 313-324, VSP, Utrecht.

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