Statistica Sinica **15**(2005), 35-57

ASYMPTOTIC PROPERTIES OF BOOTSTRAPPED LIKELIHOOD RATIO STATISTICS FOR TIME CENSORED DATA

Shuen-Lin Jeng[1], S. N. Lahiri[2] and William Q. Meeker[2]

[1] *Tunghai University and* [2] *Iowa State University*

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 time censored and the distribution of the likelihood ratio statistic is a mixture of continuous and discrete distributions. We show that the distribution of a 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 (CI) procedures based on the two parameter 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 (CBs) when the data are Type I censored.

Key words and phrases: Bartlett correction, confidence interval, life data, likelihood ratio, one-sided confidence bound, parametric bootstrap, signed square root likelihood ratio, Type I censoring.

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 (CIs) and study the accuracy of coverage probabilities for one- and two-sided CIs 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 CIs, 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) compare several CI procedures using the W_n and its Studentized modifications. Chandra and Ghosh (1979) derive 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 Nishi and Yanagimoto (1993). In two important papers, Barndorff-Nelson (1986, 1991) shows 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 arguments given by Chandra and Ghosh (1979) 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) establishes 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) and Babu and Bai (1993) establish Edgeworth expansions for statistics that are functions of lattice and non-lattice variables.

A large number of bootstrap methods have been suggested for testing or finding CIs (Hall (1992), Efron and Tibshirani (1993) and 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 CIs 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 in which the bootstrap-t method is first order accurate for 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. Proofs of the main results are given in Section 6.

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 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

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

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 $\bar{l}_n(\theta) = 1/n \sum_1^n l(x_i;\theta)$, where x_i is the data for the observation i. Let $\widehat{\theta}_n = (\widehat{\theta}_{1n}, \ldots, \widehat{\theta}_{kn})$ be the MLE of the parameter θ. Then $\widehat{\theta}_n$ satisfies the equations

$$\frac{\partial}{\partial \theta_i} \bar{l}_n(\theta) = 0, \quad i = 1, \ldots, k. \tag{2.2}$$

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 MLE of θ under the restricted model $\theta^{(2)} = \theta_0^{(2)}$. Then the log likelihood ratio statistic is

$$W_n \equiv W_n(\theta_0; k_1) = 2n[\bar{l}_n(\widehat{\theta}_n) - \bar{l}_n(\tilde{\theta}_n)] \tag{2.3}$$

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 f degree of freedom.

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

$$W_{1n} \equiv W_{1n}(\theta_0; k_1) = (k - k_1) \frac{W_n}{\mathrm{E}_{\theta_0}(W_n)} \tag{2.4}$$

and the expansion

$$\mathrm{E}_{\theta_0}(W_n) = (k - k_1) \left[1 + \frac{B(\theta_0)}{n} \right] + O\left(\frac{1}{n^2} \right). \tag{2.5}$$

Then, operationally, a Bartlett-adjusted statistic is

$$WB_n \equiv WB_n(\theta_0; k_1) = \frac{W_n}{1 + B(\tilde{\theta}^{(1)}, \theta_0^{(2)})/n}, \tag{2.6}$$

where $(\tilde{\theta}^{(1)}, \theta_0^{(2)})$ is the MLE 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

$$R_n \equiv R_n(\theta_0; k_1) = \mathrm{sign}(\widehat{\theta}_{kn} - \theta_{k0}) \sqrt{W_n}, \tag{2.7}$$

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 $\bar{\theta}_n = \widehat{\theta}_n$, 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 $f(x; \bar{\theta}_n)$. For a random variable $T_n \equiv t_n(X_1, \ldots, X_n; \theta)$, define the *parametric* bootstrap version of T_n as

$$T_n^* = t_n(X_1^*, \ldots, X_n^*; \bar{\theta}_n). \tag{2.8}$$

In absence of a parametric model, 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 consider the parametric, 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 often drop the subscript 0 and write $E_0 = E$ and $P_0 = P$. We use the following regularity conditions for proving the main results of the paper.

3.1. Conditions

We need some notation. For any real numbers x, y, 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 positive and the set of nonnegative integers. For a positive definite matrix A of order $r \in \mathbb{N}$, 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 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 $f : \mathbb{R}^k \to \mathbb{R}$, denote by $\partial^\nu f$ the partial derivative $\partial^{|\nu|}/(\partial t_1^{\nu_1}\dots\partial t_k^{\nu_k})f$ where $\nu \in \mathbb{N}^k$, $|\nu| = \sum_{i=1}^k \nu_i$ and $\nu! = \nu_1!\dots\nu_k!$. When $|\nu|=1$, we write ∂_i instead of ∂^ν to denote a partial derivative w.r.t. t_i. For a set B in $\mathbb{R}^r$, write ∂B to denote its boundary and write B^ϵ for the set $B^\epsilon = \{x \in \mathbb{R}^r : \|x-y\| \le \epsilon \quad \text{for some} \quad y \in B\}$. Also, let $e_1,\dots,e_k$ 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 $n \in \mathbb{Z}_+$,

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

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

(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

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

are non-singular, $I(\theta_n) = D(\theta_n)$ for all $n \in \mathbb{Z}_+$, and $\|I(\theta_n) - I(\theta_0)\| \to 0$ 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 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 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 $\mathcal{C}(\epsilon) = \{(t,v) : t \in \mathbb{R}^{m_1}, v \in [-\pi,\pi]^{m_2}, \|t\| \wedge \{\|v\| \vee \|p_0 - v\|\} \ge \epsilon\}$, where $p_0 = (\pi,\dots,\pi) \in \mathbb{R}^{m_2}$.

We need an 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 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 $n \in \mathbb{N}$,

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

(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 $(k_1+1,\ldots,k)$ of $A_0^{(1)} I(\theta_0)^{-1/2}$.

Condition (A.4)(ii) 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 $\tilde{Z}_{in}^{(2)}$. Babu and Bai (1993) prove a similar expansion result under a weaker moment condition than (A.4)(i), but they require a stronger Cramer condition than (A.4)(ii). Condition (A.5) ensures that the first order Taylor approximation of the target statistic depends on the continuous part $\tilde{Z}_{in}^{(1)}$, while Condition (A.6) ensures the invariance 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., (A.5) and (A.6). Of all the conditions, the uniform Cramer condition (A.4)(ii) is perhaps the most difficult to verify. A simple sufficient condition for (A.4)(ii) that, in particular, allows one to dispense with the dependence of the condition on θ_n, $n \ge 1$, is given by the following proposition.

Proposition 1. *Let $\{(X_n, Y_n)\}_{n\ge 0}$ be a collection of random vectors taking values in $\mathbb{R}^{m_1} \times \mathbb{Z}^{m_2}$. Suppose that for each $n \ge 0$, 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 $\mathbb{R}^{m_1}$ and the counting* *measure on* $\mathbb{Z}^{m_2}$ *with density* $f_n(x,y)$. *Assume that there exist a* $c>0$, *a bounded open set* $\mathcal{O} \equiv \prod_{j=1}^{m_1}(x_{0j}-a_j, x_{0j}+a_j) \subset \mathbb{R}^{m_1}$, *and integers* $l_1, l_2 \in \mathbb{Z}$ *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* $n \geq \delta^{-1}$, $\sup\Big\{\Big|E\Big[\exp(it\cdot X_n + iv\cdot Y_n)\Big]\Big| : (t,v) \in \mathcal{C}(\epsilon)\Big\} \leq (1-\delta)$.

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 $\widehat{\theta}_n$.

Theorem 1. *Assume* (A.1)−(A.3).

(a) *There exists a sequence of statistics* $\{\widehat{\theta}_n\}$ *and a constant* $a_1 \in (0,\infty)$, *independent of* n, *such that* $P_n(\|\widehat{\theta}_n - \theta_n\| \leq a_1[\log n/n]^{1/2}, \widehat{\theta}_n$ *solves* (2.2)) $= 1 - o(n^{-(s-2)/2})$.

(b) *Assume* (A.4) *and* (A.5). *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* $x \in \mathbb{R}^k$, *such that*

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

where $\Sigma_n = I(\theta_n)^{-1}$ *and* $\mathcal{B}$ *is a collection of sets in* $\mathbb{R}^k$ *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, even in the special case $\theta_n \equiv \theta_0$ for all $n \geq 1$, Theorem 1 gives an extension of the Edgeworth expansion result of Jensen (1993) for the MLE, by allowing a larger class of Borel sets than the class of convex measurable sets in $\mathbb{R}^k$; see 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 $(\sqrt{n})^{-1}\sum_{i=1}^{n}\tilde{Z}_{1n}$ 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 Götze and Hipp (1978). Although W_n admits an expansion in powers of n^{-1} in the complete data case (cf., Chandra and Ghosh (1979), Barndorff-Nielsen and Hall (1988)), 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* (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,*

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

(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 $O(n^{-2})$ (cf., Barndorff-Nielsen and Hall (1988)).

4. Results for Bootstrapped Statistics

In this section, we consider higher order accuracy of bootstrap approximations for the statistics considered in Section 2. 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 $P_{\widehat{\theta}_n}$, where $\widehat{\theta}_n$ is the MLE of θ based on $X_1, \ldots, X_n$. Let $\widehat{\theta}_n^*$ denote the bootstrap version of the MLE, obtained by replacing $X_1, \ldots, X_n$ in the definition of $\widehat{\theta}_n$ 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 $\widehat{\theta}_n$ is given by $T_n = \sqrt{n}(\widehat{\theta}_n - \theta_0)I(\widehat{\theta}_n)^{1/2}$, 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 $T_n^* = \sqrt{n}(\widehat{\theta}_n^* - \widehat{\theta}_n)I(\widehat{\theta}_n^*)^{1/2}$.

Similarly, define the bootstrap versions R_n^*, W_n^* and WB_n^* of R_n, W_n and WB_n, respectively, by replacing $X_1, \ldots, X_n$ by $X_1^*, \ldots, X_n^*$ and θ_0 by $\widehat{\theta}_n$ in (2.8), (2.4) and (2.7), respectively. For proving the results, we 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 allows us to verify the uniform Cramer condition (A.4)(ii) along different realizations of the sequence $\{\widehat{\theta}_n\}$ that lie in a set of probability 1 under θ_0.

We use modified versions of some of the Conditions (A.1)$-$(A.6). Recall that we write $E_{\theta_0} = E_0 = E$ for 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) $E_\theta\{\sup_{\|\theta - \theta_0\| \le \delta} |\partial^\nu l(X_1; \theta)\|^s\} < \delta^{-1}$ for all $\theta \in \Theta_0$.

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

- (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 $\tilde{Z}_1(\theta)$ 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.

Conditions (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). Next, write P_* for the conditional probability under $\widehat{\theta}_n$, given $X_1, \ldots, X_n$.

Theorem 3. *Suppose* (A.1)$'$−(A.5)$'$ *hold with* $s = 3$.

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

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

$$\sup_{u\in I\!R} |P_*(R_n^* \leq u) - P(R_n \leq u)| = o(n^{-1/2}) \quad a.s.(P);$$

$$\sup_{0<u<\infty} |P_*(W_n^* \leq u) - P(W_n \leq u)| = o(n^{-1}) \qquad a.s.(P);$$

$$\sup_{0<u<\infty} |P_*(WB_n^* \leq u) - P(WB_n \leq u)| = o(n^{-1}) \qquad a.s.(P).$$

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 using bootstrap versions of the SRLLR statistic R_n and the likelihood ratio statistic W_n, respectively.

5. Numerical Results

The theoretical results in this paper hold under standard regularity conditions. These conditions hold for the smallest extreme value, normal and logistic 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 incorporated 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 CIs. In Section 5.1, we describe the two parameter location-scale distribution model and in Section 5.2, we describe the simulation design and relevant formulas of the CIs. We present the results of the Weibull simulation study in Section 5.3.

5.1. The two parameter log-location-scale distribution model

We describe the general log-location-scale model that can be used for Weibull, lognormal, loglogistic and other distributions. For example, the logarithm of a Weibull random variable has a smallest extreme value distribution which is a location-scale model. Suppose that the continuous random variable $X = \log(T)$ has density $\phi_{\rm LS}[(x-\mu)/\sigma]/\sigma$ and cdf $\Phi_{\rm LS}[(x-\mu)/\sigma]$, 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

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

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

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

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. They can be checked by direct calculations for SEV, normal and logistic distributions, we omit these details here. Then for $|\nu| \le 4$,

$$\begin{aligned} Z_i = \Big(&\frac{\partial l(X_i;(\sigma,x_p))}{\partial x_p}, \frac{\partial l(X_i;(\sigma,x_p))}{\partial \sigma}, \frac{\partial^2 l(X_i;(\sigma,x_p))}{\partial x_p^2}, \frac{\partial^2 l(X_i;(\sigma,x_p))}{\partial x_p \partial \sigma}, \\ &\ldots, \frac{\partial^4 l(X_i;(\sigma,x_p))}{\partial \sigma^4}\Big), \end{aligned}$$

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. 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, (A.4)$'$ is satisfied.

The first two elements of Z_i are linearly independent when data come from the SEV, normal or logistic distribution. 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 (A.5)$'$ holds.

Because the first m_1 rows of $A^{(1)}$ give $A^{(11)}$ 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 an m_1-dimensional vector that has rank 1. Thus, (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, can be used to construct two-sided (one-sided) CIs (bounds) that are second order accurate.

5.2. The simulation design

5.2.1. Confidence intervals

This section briefly describes the different CI 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 $100(1-\alpha)\%$ CI 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 the data. For complicated problems (e.g., those involving censoring) it is necessary to estimate $\mathrm{E}(W)$ by simulation. Then an approximate $100(1-\alpha)\%$ CI 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). For estimating the scale parameter and quantiles of a positive random variable, we take the log transformation and follow the procedure in Efron and Tibshirani ((1993), Section 12.4 and 12.5).

Parametric bootstrap bias-corrected accelerated method (PBBCA). We used the procedure given by Efron and Tibshirani ((1993, Section 14.3)) who showed an easy way to obtain BC_a CIs.

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 $100(1-\alpha)\%$ CI can be computed from $\min\{R^{-1}(r_{\widehat{\theta}^*_{1(\alpha/2)}}), R^{-1}(r_{\widehat{\theta}^*_{1(1-\alpha/2)}})\}$ and $\max\{R^{-1}(r_{\widehat{\theta}^*_{1(\alpha/2)}}), R^{-1}(r_{\widehat{\theta}^*_{1(1-\alpha/2)}})\}$.

5.2.2. The simulation set up

If T has 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:

- 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 5,000 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 $p_f = 0.01$, 0.1, 0.3, 0.5, 0.9, 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 $\log(t_p)$, $p = 0.01$, 0.05, 0.1, 0.3, 0.5, 0.632 and 0.9, where $\mu \cong \log(t_{0.632})$. The one-sided $100(1-\alpha)\%$ CBs were calculated for α =0.025 and 0.05. 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 calculated results conditioned on $r > 1$.

Let $1 - \alpha$ be the nominal (user-specified) coverage probability (CP) of a procedure for constructing a CI, 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% CI from 5,000 simulations the standard error of the CP estimation is $[0.05(1-0.95)/5000]^{1/2} = 0.0031$ if the procedure is correct. The Monte Carlo error is approximately $\pm 1\%$. We say the procedure or the method for the 95% CI (or CB) 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. Because the difference of the results of the LLR and LLRB methods are not very significant, we omit the graphs of the LLR method in this paper.

Figure 1 shows the coverage probability for the one-sided approximate 95% CBs for σ from the seven methods with five different proportion failing values. Figure 2 is the same type of graph for $t_{0.1}$, the 0.1 quantile of Weibull distribution. Figure 3 shows CPs of these procedures when $p_f = 0.5$ for different quantiles. Figure 4 shows the coverage probability for 90% two-sided CIs of $t_{0.1}$. We summarize the simulation results briefly as follows.

- Using a Bartlett correction for the LLR method does not improve the coverage probability accuracy for one-sided CBs. For one-sided CBs, the LLR and LLRB methods are adequate when the expected number of failures ≥ 20. For two-sided CIs, 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 where p is close to the proportion failing, the one-sided lower CB procedure is anti-conservative. The bootstrap-t method gives accurate coverage probabilities for all functions of the parameters only 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 CBs and two-sided CIs is adequate when the number of failures exceeds 20.
- The PBSRLLR method for the one-sided CBs and two-sided CIs 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.

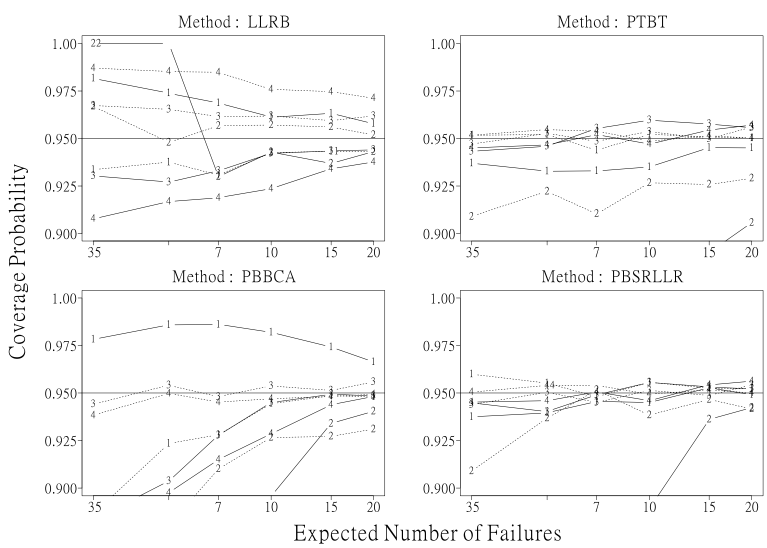

Figure 1. Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for parameter σ. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (0.01, 0.1, 0.3, 0.5, 1), respectively. Dotted and solid lines correspond to upper and lower bounds, respectively.

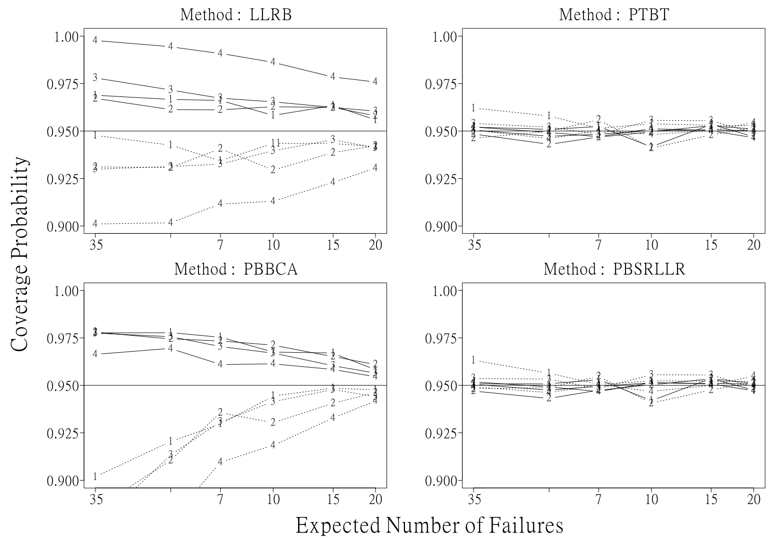

Figure 2. Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for $t_{0.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (0.01, 0.1, 0.3, 0.5, 1), respectively. Dotted and solid lines correspond to upper and lower bounds, respectively.

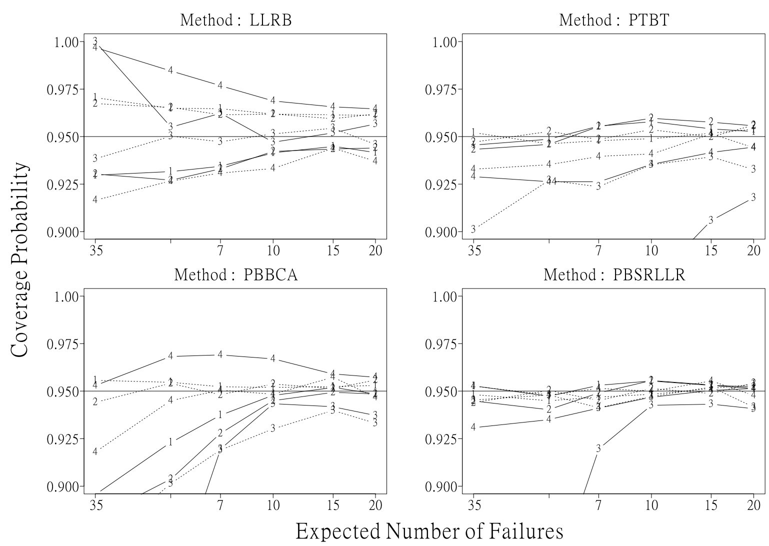

Figure 3. Coverage probability versus expected number of failures plot for one-sided approximate 95% CI procedures for $p_f = 0.5$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to t_p's, $p = (0.01, 0.1, 0.5, 0.632, 9)$, respectively. Dotted and solid lines correspond to upper and lower bounds, respectively.

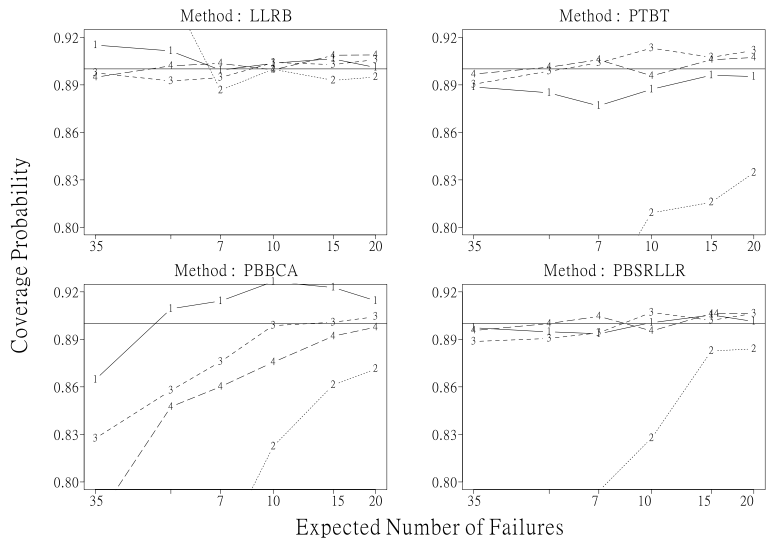

Figure 4. Coverage probability versus expected number of failures plot for two-sided approximate 90% CI procedures for $t_{0.1}$. The numbers (1, 2, 3, 4, 5) in the lines of each plot correspond to p_f's (0.01, 0.1, 0.3, 0.5, 1), respectively. Dotted and solid lines correspond to upper and lower bounds, respectively.

For Type I censored data, we can draw the following conclusions. If our interest is in constructing one-sided CBs, the PBSRLLR method provides better coverage probability with a small expected number of failures (like 10). For two-sided CIs, 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 CI 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. 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 $\mu_n(A) = \int_{A\cap B_0} f_n d\lambda, \quad A \in \mathcal{B}(\mathbb{R}^m)$. Then $\mu_n, n \geq 0$, are finite measures and by (ii), $\mu_n(\mathbb{R}^m) \to \mu_0(\mathbb{R}^m)$. Hence, by an extended version of Scheffe's Theorem (cf., Billingsley (1995, p.215)),

$$\sup\Big\{\Big|\int_{B_0} e^{i(tx+vy)} f_n(x,y)d\lambda(x,y) - \int_{B_0} e^{i(tx+vy)} f_0(x,y)d\lambda(x,y)\Big| : (t,v) \in \mathbb{R}^m\Big\}$$

$$\le \int_{B_0} |f_n - f_0| d\lambda \to 0 \text{ as } n \to \infty. \tag{6.1}$$

Fix $\epsilon > 0$, write $x_0 = (x_{01}, \ldots, x_{0m_1})$ and $m = m_1 + m_2$. Let $\varphi_1(\cdot; a)$ be the characteristic function of the UNIFORM$(-a, a)$ distribution, $a \in (0, \infty)$, and let $\varphi_2(\cdot)$ be that of the discrete uniform distribution over the integers $[l_1, l_1 + 1, \ldots, l_2]$. Then, by the property of lattice random variables (cf., Feller (1968, Chap. 15)), uniformly in $(t, v) \in \mathcal{C}(\epsilon)$,

$$\begin{aligned}
&\Big| \int_{B_0} e^{i(tx+vy)} f_0(x,y) d\lambda(x,y) \Big| \\
&\le c \Big| \sum_{y \in [l_1, l_2]^{m_2}} \Big[\int_{\mathcal{O}} e^{itx} dx \Big] e^{ivy} \Big| + \int_{B_0} [f_0(x,y) - c] d\lambda(x,y) \\
&= c \Big| \int_{\mathcal{O}} e^{it(x-x_0)} dx \Big| \cdot \prod_{j=1}^{m_2} \Big| \sum_{k=l_1}^{l_2} e^{iv_j k} \Big| + \int_{B_0} f_0(x,y) d\lambda(x,y) - c\lambda(B_0) \\
&= c\lambda(B_0) \Big\{ \prod_{j=1}^{m_1} \Big| \varphi_1(t_j; a_j) \Big| \Big\} \Big\{ \prod_{j=1}^{m_2} \Big| \varphi_2(v_j) \Big| \Big\} + \mu_0(I\!R^m) - c\lambda(B_0) \\
&< \mu_0(I\!R^m) - \epsilon_1
\end{aligned} \tag{6.2}$$

for some constant $\epsilon_1 > 0$, depending on ϵ, c, $\lambda(B_0)$. This proves the Proposition for $n = 0$. Now, the case $n = 0$ together with (6.1) gives the desired bound on the characteristic functions of (X_n, Y_n) for all large n.

Lemma 1. *Let $(X_{ni}, Y_{ni}) \in I\!R^{p+q}$, $i = 1, \ldots, n$, $n \ge 1$ be a triangular array of row i.i.d. random vectors such that for each $n \ge 1$, $Y_{n1} \in I\!R^q$ is a lattice variable with minimal lattice $Z\!\!Z^q$, $EX_{n1} = 0$, $EY_{n1} = 0$ and $Cov(X_{n1}, Y_{n1}) = I_{p+q}$. Suppose that there exists a $\delta \in (0,1)$ and an integer $s \ge 3$ such that for all $n > \delta^{-1}$,*

$$E\|X_{n1}\|^{\alpha(s)} \le \delta^{-1}, \qquad E\|Y_{n1}\|^{\beta(s)} \le \delta^{-1}, \tag{6.3}$$

for $\alpha(s) = \max\{2s+1, p+1\}$ and $\beta(s) = \max\{\alpha(s), q+1\}$, and that for any $\epsilon > 0$, there exists a $\delta \in (0,1)$ such that for all $n \ge \delta^{-1}$,

$$\sup\{|E\exp(i(tX_{n1} + vY_{n1}))| : (t,v) \in \mathcal{C}(\epsilon)\} \le 1 - \delta. \tag{6.4}$$

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 $g(0) = 0$ and $\sum_{j=1}^{p} [\partial_j g(0)]^2 = 1$. Let $S_n = ((\sqrt{n})^{-1} \sum_{i=1}^n X_{ni}, (\sqrt{n})^{-1} \sum_{i=1}^n Y_{ni})$, and let $\chi_{\nu n}$ be the νth cumulant of (X_{n1}, Y_{n1}), $\nu \in Z\!\!Z_+^{p+q}$. Then, there exist functions $p_j(\cdot;\cdot)$ such that

$$\begin{aligned}
&\sup_{u \in I\!R} \left| P\Big(\sqrt{n} g\Big(\frac{S_n}{\sqrt{n}}\Big) \le u \Big) - \int_{-\infty}^{u} \left[1 + \sum_{j=1}^{s-2} n^{-j/2} p_j(u; \{\chi_{\nu n} : |\nu| \le j+2\}) \right] \phi(u) du \right| \\
&= O(n^{-(s-1)/2}).
\end{aligned} \tag{6.5}$$

Proof. Here we outline a proof of Lemma 1 using some extensions of the arguments developed by Jensen (1989). Without loss of generality, suppose that (6.3) and (6.4) of Lemma 1 hold for all $n \geq 1$. Let $S_n^{(1)} = (\sqrt{n})^{-1} \sum_{i=1}^n X_{ni}$ and $S_n^{(2)} = (\sqrt{n})^{-1} \sum_{i=1}^n Y_{ni}$. By Bartlett's (1938) formula for the conditional characteristic function,

$$E[\exp(itS_n^{(1)})|S_n^{(2)} = y] = (2\pi)^{-q} \int_{[-\pi\sqrt{n},\pi\sqrt{n}]^q} \left[f_n^n\left(\frac{t}{\sqrt{n}}, \frac{v}{\sqrt{n}}\right)\right] e^{-ivy} dv/a_n, \tag{6.6}$$

where $a_n = n^{q/2} P(S_n^{(2)} = y)$ 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

$$\psi_n(t, v) = e^{-(\|t\|^2+\|v\|^2)/2}\Big[1 + \sum_{j=1}^{\alpha(s)-3} n^{-j/2}\tilde{P}_j(it, iv; \{\chi_{\nu n} : |\nu| \leq j + 2\})\Big], \tag{6.7}$$

where the functions $\tilde{P}_j(it, iv; \{\chi_{\nu n} : |\nu| \leq j + 2\})$, $j \geq 1$, are defined by identity (7.2) of [BR] (with χ_ν's there replaced by $\chi_{\nu n}$'s). Also, let $\widehat{\Psi}_n(t|y)$ be defined by

$$\widehat{\Psi}_n(t|y) = (2\pi)^{-q} \int \psi_n(t, v)e^{-ivy} dv/b_n, \tag{6.8}$$

where $b_n \equiv (2\pi)^{-q} \int \psi_n(0, v)e^{-ivy} du$. By (6.3) and arguments in (2.9)–(2.12) of Jensen (1989), it follows that, uniformly in $\|y\|^2 \leq s\log n$, $|a_n - b_n| = O(n^{-(\alpha(s)-2)/2})$ and $a_n \wedge b_n \geq C_1 n^{-s/2}$ for some $C_1 \in (0, \infty)$, so that

$$|a_n - b_n|a_n^{-1} = O(n^{-(\alpha(s)-2-s)/2}) = O(n^{-(s-1)/2}). \tag{6.9}$$

Arguments leading to Theorem 9.9 of [BR] yield

$$\begin{aligned} &\left|\partial^\nu(f_n^n(\frac{t}{\sqrt{n}}, \frac{v}{\sqrt{n}}) - \psi_n(t, v))\right| \\ &\leq C_2(\delta)n^{-(\alpha(s)-2)/2}\left[1 + \|t\|^{3(\alpha(s)-2)+|\nu|}\right] e^{-(\|t\|^2+\|u\|^2)/4} \end{aligned} \tag{6.10}$$

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

$$\sup_{B\in\mathcal{B}}\left|P(S_n^{(1)} \in B|S_n^{(2)} = y) - \Psi_n(B|y)\right| = O(n^{-(s-1)/2}) \tag{6.11}$$

uniformly over $\sqrt{n}y \in Z\!\!\!Z^{m_2}$ with $\|y\|^2 \leq s\log n$, where $\Psi(\cdot|y)$ is the signed measure corresponding to the Fourier transform $\widehat{\Psi}(\cdot|y)$ of (6.8), and $\mathcal{B}$ is any given collection of Borel subsets of $I\!R^p$ satisfying '$\Phi([\partial B]^\epsilon) = O(\epsilon)$ as $\epsilon \to 0$'.

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

$$\sup_{u_0 \in I\!R} \Big| P\Big(\sqrt{n} g\Big(\frac{S_n}{\sqrt{n}}\Big) \leq u_0 | S_n^{(2)} = y\Big) - \int_{-\infty}^{u_0} \Big[1 + \sum_{j=1}^{s-2} n^{-1/2} \check{p}_{jn}(u; y)\Big] \phi_{\sigma_n}(u - d_n y) du \Big| = O(n^{-(s-1)/2}) \quad (6.12)$$

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 Götze and Hipp (1978) and an analog of (6.10) with $t = 0, |\nu| \leq \beta(s)$, one gets (6.5).

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

$$0 = \partial_j \bar{l}_n(\theta) = \partial_j \bar{l}_n(\theta_n) + \sum_{|\nu|=1}^{s-1} [\partial^\nu \partial_j \bar{l}_n(\theta_n)](\theta - \theta_n)^\nu / \nu! + R_{nj}, \quad (6.13)$$

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

Using (A.2), (A.3), and Corollary 4.2 of Fuk and Nagaev (1971), we get $P_n(|\partial^\nu \bar{l}_n(\theta_n) - E_n \partial^\nu \bar{l}_n(\theta_n)| > C n^{-1/2} (\log n)^{1/2}) = O(n^{-(s-2)/2} (\log n)^{-s/2})$, $1 \leq |\nu| \leq s-1$, and $P_n\left(\sup\{|\partial^\nu \bar{l}_n(\theta)| : \|\theta - \theta_n\| \leq a_1, |\nu| = s+1\} > C\right) = O(n^{-(s-2)/2} (\log n)^{-s/2})$. Hence, on a set A_n with $P_n(A_n^c) = O(n^{-(s-2)/2} (\log n)^{-s/2})$, we may rewrite (6.13) as

$$(\theta - \theta_n) = g_n(\theta - \theta_n) \quad (6.14)$$

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 $\widehat{\theta}_n$ and θ_n as

$$\widehat{\theta}_n = g(\bar{Z}_n^\dagger) \text{ and } \theta_n = g(E_n \bar{Z}_n) \quad (6.15)$$

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 $Z_{in}^{\dagger(\nu)} = Z_{in}^{(\nu)} + R_n(\widehat{\theta}_n)$, $|\nu| = 1$, and $R_n(\widehat{\theta}_n)$ is the vector of $R_{nj}(\widehat{\theta}_n)$, $1 \leq j \leq k$.

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). We omit the routine details.

Proof of Theorem 2. Following the arguments on pp.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 $m = 1, 2$,

$$\tilde{V}_{mn} = \sum_{i=k_1+1}^{k} a_{mi}\Big[\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\tilde{Z}_{in}^{(e_i)}\Big] + \sum_{r=1}^{2} n^{-r/2}\check{p}_{rm}\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\tilde{Z}_{in};\theta_n\Big) + \check{R}_{mn}, \quad (6.16)$$

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 $m = 1$). Here, the remainder term $\check{R}_{mn}$ satisfies the inequality $P_n(|\check{R}_{mn}| > Cn^{-1}(\log n)^{-2}) = O(n^{-1}(\log n)^{-2})$ for $m = 1, 2$. Now applying Lemma 1 above and the transformation techniques of [BG], one can establish (3.8) and (3.9) of Theorem 2. The proof of (3.10) is similar, by noting that the effect of the correction factor $1/(1 + B(\cdot))$ 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 $s = 3$, $E_0\|Z_{10}\|^{2s+1} \le \|A_0\|^{2s+1}E_0\|\tilde{Z}_{10}\|^{2s+1} < \infty$, so that

$$P_0(\|\widehat{\theta}_n - \theta_0\| > a_1[(\log n)/n]^{1/2}) = o(n^{-5/2}). \quad (6.17)$$

Hence, by the Borel-Cantelli lemma, $\widehat{\theta}_n - \theta_0 = O(n^{-1/2}(\log n)^{1/2})$ a.s. (P_0). Let D_0 be the set of P_0-probability 1 where $\widehat{\theta}_n - \theta_0 = O(n^{-1/2}(\log n)^{1/2})$ as $n \to \infty$. Then, by the continuity of $\partial^\nu l(x;\theta)$ in θ over Θ, $1 \le |\nu| \le s$, and the continuity of the second moments, $f(x, y; \widehat{\theta}_n) \to f(x, y; \theta_0)$ 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 $\{\widehat{\theta}_n\}$ on D_0. Now, using the expansion for $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ from part (b) of Theorem 1 and the transformation technique of [BG], one can show

$$\sup_{B\in\mathcal{B}}\Big|P(T_n^* \in B) - \int_B \Big[1 + q_1(x;\widehat{\theta}_n)n^{-1/2}\Big]\phi(x)dx\Big| = o(n^{-1/2}) \quad a.s.(P_0), (6.18)$$

$$\sup_{B\in\mathcal{B}}\Big|P(T_n \in B) - \int_B \Big[1 + q_1(x,\theta_0)n^{-1/2}\Big]\phi(x)dx\Big| = o(n^{-1/2}), \quad (6.19)$$

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.

Acknowledgement

This research was partially supported by NSF grant number DMS 0072571.

References

Babu, G. J. (1991). Edgeworth expansions for statistics which are functions of lattice and non-lattice variables. *Statist. Probab. Lett.* **12**, 1-7.

Babu, G. J. and Bai, Z. D. (1993). Edgeworth expansion of a function of sample means under minimal moment conditions and partial Cramer's condition. *Sankhyā A* **55**, 244-258

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 Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75**, 374-378.

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

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

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

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

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 statistic and other perturbed chi-square variables. *Sankhyā A* **41**, 22-47.

Datta, S. (1992). A note on continuous 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.* Chapman and Hall, New York.

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

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

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

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

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

Horvath, L. and Yandell, B. S. (1987). Convergence rates for bootstrapped product-limit process. *Ann. Statist.* **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. *Ann. Statist.* **15**, 314-324.

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

Jensen, J. L. (1993). A historical sketch and some new results on the improved log likelihood ratio statistic. *Scand. J. Statist.* **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. *Statist. Sinica* **3**, 517-542.

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

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

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

Nishi, R. and Yanagimoto, T. (1993). Normal approximation to the distribution of the sample mean in the exponential family. *Statist. Sci. Data Analysis*, 313-324. VSP, Utrecht.

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

Department of Statistics, Tunghai University, Taichung, 407, Taiwan, R.O.C.

E-mail: sljeng@mail.thu.edu.tw

Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.

E-mail: snlahiri@iastate.edu

Department of Statistics, Iowa State University, Ames, IA 50011, U.S.A.

E-mail: wqmeeker@iastate.edu

(Received February 2003; accepted August 2004)