

# Simultaneous Parametric Confidence Bands for Cumulative Distributions from Censored Data

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

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

February 18, 2000

## Abstract

This paper describes existing methods and develops new methods for constructing simultaneous confidence bands for a cumulative distribution function (cdf). Our results are built on extensions of previous work by Cheng and Iles (1983, 1988). Cheng and Iles use Wald statistics with (expected) Fisher information and provide different approaches to find one-sided and two-sided simultaneous confidence bands. We consider three statistics, Wald statistics with Fisher information, Wald statistics with local information, and likelihood ratio statistics. Unlike pointwise confidence intervals, it is not possible to combine two 95% one-sided simultaneous confidence bands to get a 90% two-sided simultaneous confidence band. We present a general approach for construction of two-sided simultaneous confidence bands on a cdf for a continuous parametric model from complete and censored data. Both two-sided and one-sided simultaneous confidence bands for the location-scale parameter model are discussed in detail including situations with complete and censored data. We start by using standard large-sample approximations and then extend and compare these to corresponding simulation or bootstrap calibrated versions of the same methods. The results show that bootstrap methods provide more accurate coverage probabilities than those based on the usual large sample approximations. A simulation for the Weibull distribution and Type I censored data is used to compare finite sample properties. For the location-scale model with complete or Type II censoring, the bootstrap methods are exact. Simulation results show that, with Type I censoring, a bootstrap method based on the Wald statistic with local information provides a confidence region with coverage probabilities that are more accurate than a method based on bootstrapping the likelihood ratio statistic. We illustrate the implementation of the methods with an application to estimate probability of detection (POD), a function that is used to assess nondestructive evaluation (NDE) capability.

**Keywords:** Bootstrap, likelihood ratio, simultaneous confidence band, life data, probability of detection, Wald.

# 1 Introduction

## 1.1 Problem

In life testing and reliability studies, the primary problem of interest is often to estimate an unknown cumulative distribution function (cdf). For example, sample units might be put on a life test for the purpose of estimating the proportion failing before some specific time point. Another example is the need to quantify nondestructive evaluation (NDE) capability. NDE methods are used, for example, to detect subsurface flaws before processing expensive materials. Inputs for a risk analysis include detection capability for a range of different flaw sizes. These problems can be formulated as one where an unknown cdf is to be estimated. We will, however, use the more familiar failure time language in our general discussion.

Confidence intervals quantify the uncertainty of estimation. For example, pointwise confidence intervals with a specific confidence level can be computed for the cdf at particular times. When the interest is on the cdf over a range of time values, the procedure using the combination of these pointwise confidence intervals will not provide a simultaneous confidence band with same coverage probability. For a given confidence level, a simultaneous confidence band would be wider than the joint set of pointwise confidence intervals. This is because we use the same information from the data to do the inference for specific point of interest as we have for inference on an infinite number of points.

Unlike pointwise confidence intervals, one cannot combine two  $100(1 - \alpha/2)\%$  one-sided simultaneous confidence bands to get a  $100(1 - \alpha)\%$  two-sided simultaneous confidence band. Somewhat different procedures are needed for one-sided and two-sided cases.

Censoring often arises in life data collection. Some theoretical results for complete data do not hold for censored data. Especially for Type I censoring, the Wald and the likelihood ratio statistics no longer have the pivotal property (a pivotal statistic has a distribution that does not depend on unknown parameters) in location-scale models. Bootstrap methods, however, provide a more accurate approximate distribution when the exact distributional form is not available. Jeng and Meeker (1999) show that the bootstrap likelihood ratio procedures are generally second order accurate for complete and censored data. Simulation results in this paper show that the procedure based on the bootstrap Wald statistics with local information provides a confidence region with a confidence level that appears to be as accurate as or more accurate than the procedure based on the bootstrap likelihood ratio statistics, even when the expected number of failures is small.

## 1.2 Literature Review

Nonparametric methods for constructing confidence bands for cdfs can, for example, be based on statistics like the Kolmogorov-Smirnov statistics. See Lehmann (1986, pp. 355-357) for definition and references to the literature. As described in Cheng and Iles (1983), however, these methods give rise to a constant (vertical) width and part of such a band will have ordinate values that are greater than one, while other parts will have ordinate values that are negative. Even if the general approach is used in a parametric setting, it makes the band unnecessarily broad in the tails. Kanofsky and Srinivasan (1972) overcome the problem under normal, exponential and uniform models by using the maximum absolute difference between the true function and an estimator of it (similar to the Kolmogorov-

Smirnov statistics) and by adjusting the resulting band to obtain the required confidence level. Using the Wald statistics with (expected) Fisher information, Cheng and Iles (1983) provide an alternative general procedure that can be applied to construct simultaneous bands for any continuous function  $g(\cdot; \theta)$  of the parameters  $\theta$ . First, a joint confidence region is constructed for the unknown parameters. Then a simultaneous confidence band is obtained by seeing how the continuous function  $g(\cdot; \theta)$  changes as the parameters are varied within the joint confidence region. The band is two-sided and has ordinate values that lie within the range of  $g$ . Cheng and Iles (1988) extend the result to one-sided simultaneous confidence bands for a cdf under the location-scale model with complete data. The simultaneous confidence bands constructed in this way may be exact or conservative. We show that, for the location-scale model with complete or Type II censoring, the bootstrap methods we present for constructing the two-sided and one-sided simultaneous confidence bands are exact.

Le Cam (1990) replaced Wald's confidence ellipsoids by confidence sets based on Hellinger distance. Escobar and Meeker (2000) develop methods parallel to Cheng and Iles (1983) based on the Wald statistics with local information for two-sided bands of quantiles and cdfs and show a certain duality property.

A likelihood ratio test can be used to construct a joint confidence region (or an approximate joint confidence region) for model parameters. The likelihood ratio statistics are transformation invariant, unlike the Wald statistics. Generally the distribution of the likelihood ratio statistic follows a  $\chi^2$  distribution to the order  $O(1/n)$  for both complete and censored data (Jensen, 1993). This likelihood confidence region can produce simultaneous confidence bands for cdf's or any continuous function  $g(\cdot; \theta)$ . Satten (1995) uses likelihood confidence region to construct simultaneous confidence bands for quantiles from iid samples.

In the location-scale model for complete and Type II censored data, the Wald statistic is a pivotal statistic. One can find the distribution of the Wald statistic by using simulation (or parametric bootstrap) methods. For time-censored data, the distribution of the Wald and the likelihood ratio statistics depends on the unknown proportion in the population that would fail before the fixed censoring time. The bootstrap procedure still provides a second order accurate approximation for the distribution of likelihood ratio statistics (see Jeng and Meeker, 1999).

There are some other bootstrap methods for constructing joint confidence regions that are not included in this research. Beran (1988) suggests a method called bootstrap pre-pivoting to find the simultaneous confidence bands for a family of parametric functions. The advantage of Beran's method is that the resulting confidence intervals are asymptotically balanced. A simultaneous confidence band of a function  $g(\cdot, \theta)$  is balanced if the pointwise confidence level for the confidence statement concerning  $g(x, \theta)$  remains unchanged as  $x$  varies. But the pre-pivoting procedure usually needs a double bootstrap to make the root closer to a pivot. Hall (1992, Section 4.2) suggests a likelihood based region that requires high dimensional density estimation. Yeh and Singh (1997) propose a bootstrap balanced confidence region based on the Tukey depth. The difficulty of using this method is the large amount of computer time required to calculate the Tukey depth accurately.

Simultaneous confidence bands for POD function are important in aircraft life management. Annis et al. (MIL-STD-1823, 1989) describe a normal-approximation based lower simultaneous confidence bound for POD function, using an extension of the Cheng and

Iles (1983) method. Meeker et al. (1995, 1996, 1997) develop a methodology to estimate Nondestructive Evaluation (NDE) capability. The methodology is based on a physical/statistical model and can be used to estimate probability of detection (POD) curves and other characteristics of a flaw-detecting system. Sarkar et al. (1998) apply a similar method to quantify nondestructive testing inspection capability, using data from destructive testing of cracks found in heat exchanger tubes. Their data were right censored because of measurement saturation for large signals. They estimate a POD curve for a particular flaw detection system and provide pointwise confidence intervals for the POD curve based on the delta method and a normal approximation. We extend the results to provide simultaneous confidence bands for the POD curve. We use a bootstrap procedure to build a joint confidence region for the unknown parameters. This bootstrap procedure is similar to the one used by Robinson (1983) to construct confidence intervals for one-dimensional parameters from progressively censored data. Then the joint confidence region is used to construct a simultaneous confidence band for the POD curve.

### 1.3 Overview

Section 2 provides a general approach for constructing two-sided simultaneous confidence bands for a function  $g(\cdot; \theta)$ . Section 3 focuses on the location-scale distribution model and Section 4 presents the results of a simulation study using the Weibull distribution with complete and Type I censored data. Section 5 presents an application in which the simultaneous confidence bands are used to quantify the uncertainty in the probability of detection curve. Section 6 provides a summary of our results and outlines possible areas for future research.

## 2 Methods

Let  $W$  be a continuous random variable with cdf  $F(w; \theta)$  and let  $g(x; \theta)$  be a continuous function defined on a set  $\mathcal{D}$  (e.g., the positive real line) and a  $k$  dimensional parameter space of  $\theta$ . For example  $g(x; \theta)$  could be the cdf of  $W$ . A random sample  $w_1, \dots, w_n$  of size  $n$  is to be used to calculate a simultaneous confidence band for  $g(x; \theta)$  over some specified (possibly infinite) range of  $x$  values. We present a general approach for constructing two-sided simultaneous confidence bands. The method can be used for both complete and censored data. The approach extends previous results from Cheng and Iles (1983).

Suppose first that  $\mathcal{R}$  is a  $100(1 - \alpha)\%$  joint confidence region for the unknown parameter vector  $\theta$ .  $\mathcal{R}$  could be obtained for the purpose of constructing either one-sided or two-sided simultaneous confidence bands. For a given function  $g$ , let us consider the function  $y = g(x; \theta)$  in the  $(x, y)$  plane for  $x \in \mathcal{D}$ . When  $\theta$  is changing in  $\mathcal{R}$ , the function  $g$  will cover a region,  $\mathcal{B}$ , on the  $(x, y)$  plane. Because the true value of  $\theta$  lies in  $\mathcal{R}$  with probability  $1 - \alpha$ , the probability is at least  $1 - \alpha$  that one of the functions used to cover the region  $\mathcal{B}$  is the unknown true function  $g(\cdot; \theta)$ . Thus  $\mathcal{B}$  is a simultaneous confidence band for  $g(\cdot; \theta)$  that will contain the true function  $g(\cdot; \theta)$  with probability at least  $1 - \alpha$ . In general there may be values of  $\theta$  outside of the region  $\mathcal{R}$  that give a function  $g(\cdot; \theta)$  lying entirely within the band  $\mathcal{B}$ . So the band  $\mathcal{B}$  could be conservative.

Define the lower and upper confidence curves  $C_l$  and  $C_u$  at  $x$  corresponding to a joint confidence region  $\mathcal{R}$  as

$$C_l(x) = \min_{\theta \in \mathcal{R}} g(x; \theta), \quad C_u(x) = \max_{\theta \in \mathcal{R}} g(x; \theta). \quad (1)$$

If  $\mathcal{R}$  is the region constructed for a two-sided simultaneous confidence band, we denote the two-sided band by

$$\mathcal{B} = \{(x, y) : C_l(x) \leq y \leq C_u(x), x \in \mathcal{D}\}. \quad (2)$$

Usually in order to achieve the required confidence level, a different joint confidence region  $\mathcal{R}$  is needed for a one-sided simultaneous band. For a region  $\mathcal{R}$  constructed to compute a one-sided band, we use the region to produce a lower confidence curve  $C_l(x)$  and denote the one-sided lower simultaneous confidence band by

$$\mathcal{B}_l = \{(x, y) : y \geq C_l(x), x \in \mathcal{D}\}. \quad (3)$$

Similarly, we denote the one-sided upper simultaneous confidence band by

$$\mathcal{B}_u = \{(x, y) : y \leq C_u(x), x \in \mathcal{D}\}. \quad (4)$$

## 2.1 Methods Used

The different methods for constructing a joint confidence region  $\mathcal{R}$  are based on different statistics and procedures. Below we describe briefly seven methods by indicating how the exact or approximate distribution of the statistics are obtained. In all of these methods, we let  $L(\theta)$  denote the likelihood function and  $\hat{\theta}$  denote the maximum likelihood estimator of  $\theta$  and assume that the standard set of regularity conditions holds.

### 2.1.1 $\chi^2$ -approximation Methods

**Wald statistic with Fisher information (WLADF).** Let

$$I(\theta) = \text{E} \left[ \frac{\partial \log L(\theta)}{\partial \theta_i} \frac{\partial \log L(\theta)}{\partial \theta_j} \right]$$

be the Fisher (expected) information matrix. The Wald statistic with Fisher information is

$$Q_F(\theta) = (\hat{\theta} - \theta)' I(\theta) (\hat{\theta} - \theta).$$

Rao (1973, page 418) shows that the large-sample limiting distribution of  $Q_F$  is  $\chi_k^2$ .

**Wald statistic with local information (WLADL).** Let

$$\hat{I}(\hat{\theta}) = \left[ \frac{\partial \log L(\hat{\theta})}{\partial \theta_i} \frac{\partial \log L(\hat{\theta})}{\partial \theta_j} \right]$$

be the local information matrix. The Wald statistic with local information is

$$Q_L(\theta) = (\hat{\theta} - \theta)' \hat{I}(\hat{\theta}) (\hat{\theta} - \theta).$$

Cox and Hinkley (1974, page 314) show that the large-sample limiting distribution of  $Q_L$  is  $\chi_k^2$ .

**Log LR method (LLR).** The likelihood ratio statistic is defined as

$$W(\theta) = -2 \log \left[ \frac{L(\theta)}{L(\hat{\theta})} \right].$$

Serfling (1980, Section 4.4) shows that the large-sample limiting distribution of  $W(\theta)$  is  $\chi_k^2$ .

**Log LR Bartlett corrected method (LLRB).** The Bartlett corrected likelihood ratio statistic is

$$W_B(\theta) = k \frac{W(\theta)}{E[W(\theta)]}.$$

Because the expectation of  $W_B(\theta)$  is equal to the mean of the  $\chi_k^2$  distribution, the distribution of  $W_B(\theta)$ , when compared with  $W(\theta)$ , can be expected to be better approximated by the  $\chi_k^2$  distribution (Bartlett 1937). In general one must substitute an estimate for  $E[W(\theta)]$  computed from one's data. For complicated problems (e.g., those involving censoring) it is necessary to estimate of  $E[W(\theta)]$  by using simulation.

### 2.1.2 Parametric Bootstrap Methods

The following methods use the “bootstrap principle” or Monte Carlo evaluation of sampling distributions. Suppose a statistic  $S(\theta)$  is a function of random variables with a distribution that depends on the parameter  $\theta$ . The parametric bootstrap version  $S^*(\hat{\theta})$  of  $S$  is the same function but evaluated at data (“bootstrap samples”) simulated using an estimate  $\hat{\theta}$  instead of the unknown  $\theta$  [see Sec. 6.5, Efron and Tibshirani (1993) for more details]. Using  $\hat{\theta}$  in place of the distribution parameters, the distribution of  $S^*$  can be calculated analytically in simple situations, or by simulation in general. Except for special cases in which the underlying statistic is pivotal [e.g., complete data or Type II censoring from location-scale distributions] the distribution of  $S^*$  will depend on  $\hat{\theta}$ , and thus the distribution of  $S^*$  will provide only an approximation to the distribution of  $S$ .

**Parametric bootstrap Wald statistic with Fisher information (BWALDF).** Let  $Q_F^*(\theta)$  be the bootstrap version of  $Q_F(\theta)$ . Use the distribution of  $Q_F^*(\theta)$  to approximate the distribution of  $Q_F(\theta)$ .

**Parametric bootstrap Wald statistic with local information (BWALDL).** Let  $Q_L^*(\theta)$  be the bootstrap version of  $Q_L(\theta)$ . Use the distribution of  $Q_L^*(\theta)$  to approximate the distribution of  $Q_L(\theta)$ .

**Parametric bootstrap log likelihood ratio method (BLLR).** Let  $W^*(\theta)$  be the bootstrap version of  $W(\theta)$ . Use the distribution of  $W^*(\theta)$  to approximate the distribution of  $W(\theta)$ .

## 2.2 Construction of Simultaneous Confidence Bands

Let  $S(\theta)$  be any one of  $Q_L(\theta)$ ,  $Q_F(\theta)$ ,  $W(\theta)$ , or  $W_B(\theta)$ . Also let  $\gamma$  denote the  $100(1 - \alpha)\%$  quantile from the distributions corresponding to  $S(\theta)$  or  $S^*(\theta)$  from one of the seven

methods. A  $100(1 - \alpha)\%$  confidence region for  $\theta$  can be obtained by

$$\mathcal{R} = \{\theta : S(\theta) < \gamma\}. \quad (5)$$

By using the notation of (1), an approximate  $100(1 - \alpha)\%$  two-sided simultaneous confidence band can be obtained from (2).

In the next section we show that under the location-scale model with complete or Type II censored data, the coverage probability of the bootstrap methods for constructing two-sided simultaneous confidence bands for the cdf by using the joint confidence region  $\mathcal{R}$  of  $\theta$  is equal to the nominal confidence level. When the data are Type I censored from a location-scale model or when the data are complete from other general models, approximate two-sided simultaneous confidence bands still can be constructed using these methods.

The confidence level for a one-sided confidence band constructed by using the joint confidence region obtained with equation (5) will be larger than that for the associated region. In general the joint confidence region needed for the one-sided simultaneous confidence bands depends on the properties of function  $g$ . However, for the location-scale model with complete data, Cheng and Iles (1988) give a procedure that provides a one-sided bound with the correct coverage probability.

Jeng and Meeker (1999) show that under some regularity conditions the BLLR method is second order accurate even for Type I censored data. This procedure, when used to construct joint confidence regions, has a coverage probability that is close to nominal. The BLLR method also provides approximate two-sided and one-sided simultaneous confidence bands for general models with complete or censored data.

In the next section we focus on methods for the location-scale model with complete or censored data. These methods can, however, be applied to log-location-scale distributions like the lognormal, Weibull and loglogistic.

### 3 Location-scale Model

Suppose  $\Phi(\xi)$  is a known continuous distribution function, and consider a random variable  $X$  with cdf  $\Phi[(x - \mu)/\sigma]$  and density  $\phi[(x - \mu)/\sigma]/\sigma$  where  $\mu$  and  $\sigma$  are the unknown location and scale parameters. In this case  $X$  is said to have a location-scale distribution. Let  $\hat{\mu}$  and  $\hat{\sigma}$  be the maximum likelihood estimators for  $\mu$  and  $\sigma$ .

#### 3.1 Two-sided Simultaneous Confidence Bands

This section describes some properties of the statistics that are used to construct joint confidence regions. First we express these statistics into different forms.

**WALDF.** The Fisher information matrix for  $\mu$  and  $\sigma$  can be written as

$$I(\mu, \sigma) = \frac{n}{\sigma^2} \begin{pmatrix} i_{11} & -i_{12} \\ -i_{12} & i_{22} \end{pmatrix}.$$

Then, as shown in Cheng and Iles (1988), the Wald statistic with Fisher information can

be expressed as

$$\begin{aligned} Q_F(\mu, \sigma) &= ni_{11}(\hat{\mu} - \mu)^2/\sigma^2 - 2ni_{12}(\hat{\mu} - \mu)(\hat{\sigma} - \sigma)/\sigma^2 + ni_{22}(\hat{\sigma} - \sigma)^2/\sigma^2 \\ &= ni_{11}M^2 - 2ni_{12}MS + ni_{22}S^2, \end{aligned}$$

where  $M = (\hat{\mu} - \mu)/\sigma$  and  $S = (\hat{\sigma} - \sigma)/\sigma$ .

**WALDL.** Similarly the local information matrix can be written as

$$\hat{I} = I(\hat{\mu}, \hat{\sigma}) = \frac{n}{\hat{\sigma}^2} \begin{pmatrix} \hat{i}_{11} & -\hat{i}_{12} \\ -\hat{i}_{12} & \hat{i}_{22} \end{pmatrix}.$$

Then the Wald statistic with local information can be expressed as

$$Q_L(\mu, \sigma) = n\hat{i}_{11}\hat{M}^2 - 2n\hat{i}_{12}\hat{M}\hat{S} + n\hat{i}_{22}\hat{S}^2,$$

where  $\hat{M} = (\hat{\mu} - \mu)/\hat{\sigma}$  and  $\hat{S} = (\hat{\sigma} - \sigma)/\hat{\sigma}$ .

**LLR.** For complete data, the likelihood ratio statistic can be expressed as

$$W(\mu, \sigma) = -2 \log \left[ \left[ \left( \frac{\hat{\sigma}}{\sigma} \right)^n \right] \frac{\prod_{i=1}^n \phi \left( \frac{x_i - \mu}{\sigma} \right)}{\prod_{i=1}^n \phi \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)} \right].$$

For right censored data (Type I or Type II), let  $\delta_i = 1$  if the  $i$ th observation is a failure,  $\delta_i = 0$  if  $i$ th the observation is censored. Then the likelihood ratio statistics can be expressed as

$$W(\mu, \sigma) = -2 \log \left\{ \left[ \left( \frac{\hat{\sigma}}{\sigma} \right)^{\sum_{i=1}^n \delta_i} \right] \frac{\prod_{i=1}^n \phi \left( \frac{x_i - \mu}{\sigma} \right)^{\delta_i} \left[ 1 - \Phi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{1 - \delta_i}}{\prod_{i=1}^n \phi \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)^{\delta_i} \left[ 1 - \Phi \left( \frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \right]^{1 - \delta_i}} \right\}.$$

**Complete data.** Kendall and Stuart (1991) show that  $i_{11}, i_{12}$ , and  $i_{22}$  are constants independent of  $\mu$  and  $\sigma$ . Because the distributions of  $M$  and  $S$  do not depend on  $\mu$  and  $\sigma$  (Lawless 1982, page 147),  $Q_F$  is a pivotal quantity. Note that  $(x_i - \hat{\mu})/\hat{\sigma}$ ,  $\hat{M}$ , and  $\hat{S}$  are functions of  $\hat{\sigma}/\sigma$ ,  $(x_i - \mu)/\sigma$ ,  $M$ , and  $S$ , so the distribution of  $(x_i - \hat{\mu})/\hat{\sigma}$ ,  $\hat{M}$ , and  $\hat{S}$  do not depend on  $\mu$  or  $\sigma$ . The elements  $\hat{i}_{11}, \hat{i}_{12}$ , and  $\hat{i}_{22}$  depend only on  $(x_i - \hat{\mu})/\hat{\sigma}$ . Thus  $Q_L$  is also a pivotal quantity. Because  $W$  depends only on  $(x_i - \mu)/\sigma$ ,  $(x_i - \hat{\mu})/\hat{\sigma}$ , and  $\hat{\sigma}/\sigma$ , it is also a pivotal quantity. The BWALDL, BWALDF, and BLLR methods for constructing the  $100(1 - \alpha)\%$  confidence regions have exact confidence level  $1 - \alpha$  except for the Monte Carlo simulation error (which can be made arbitrary small by increasing the number of Monte Carlo trails).

**Type II censored data.** Lawless (1982, page 147) shows that, with Type II censoring,  $Z_1 = (\mu - \hat{\mu})/\hat{\sigma}$ ,  $Z_2 = \hat{\sigma}/\sigma$ ,  $Z_3 = (\mu - \hat{\mu})/\sigma$ , and  $a_i = (x_i - \hat{\mu})/\sigma$  are pivotal quantities. Because  $Q_F$ ,  $Q_L$ , and  $W$  only depend on  $Z_1$ ,  $Z_2$ ,  $Z_3$ , and  $a_i$ , they are also pivotal quantities with Type II censoring. The  $100(1 - \alpha)\%$  confidence regions obtained by the BWALDL,

BWALDF, and BLLR methods have exact coverage probability  $1 - \alpha$  except for Monte Carlo simulation error.

**Type I censored data.** With Type I censoring the distributions of  $Q_F$ ,  $Q_L$ ,  $W$ , and  $W_B$  depend on the unknown proportion failing at the censoring time. For this reason, joint confidence regions and simultaneous confidence bands based on these statistics are only approximate. The approximation improves with increasing sample size.

Once we have a  $100(1 - \alpha)\%$  confidence region  $\mathcal{R}$  for  $\mu$  and  $\sigma$  from one of the previous described methods, the  $100(1 - \alpha)\%$  two-sided simultaneous confidence curves can be obtained by using the equation (1), providing the simultaneous band as indicated in (2).

Cheng and Iles (1983) show that the confidence level of the two-sided simultaneous confidence band is the same for the location-scale model as the confidence level of the confidence region produced by the WALDF method. We extend this result to show that any convex confidence region with required confidence level can be used to construct a two-sided simultaneous band and that both the region and the band have the required confidence level. Note that the confidence regions for the two-sided simultaneous bands constructed from the WALDF or the WALDL methods are ellipses and thus are convex. The LLR method will produce convex confidence regions for the log-location-scale model, but not in general.

We first use a convex confidence region  $\mathcal{R}$  to construct a simultaneous confidence band for quantiles of the distribution. Then we show that the band can be converted to a simultaneous confidence band for the cdf and argue that in either case, the confidence level of the band is the same as the confidence region  $\mathcal{R}$ .

The  $p$  quantile  $x_p$  is defined as

$$x_p = \mu + u_p\sigma, \quad (6)$$

where  $u_p = \Phi^{-1}(p)$ . Consider a fixed  $p$ ,  $0 < p < 1$ . In the  $(\mu, \sigma)$  plane, equation (6) represents a family of parallel lines with different intercepts  $x_p$  and the same slope  $-u_p^{-1}$ . Because the region  $\mathcal{R}$  is convex, the smallest and the largest values of  $x_p$  produced by  $(\mu, \sigma) \in \mathcal{R}$ , say  $\hat{x}_p(\min)$  and  $\hat{x}_p(\max)$ , correspond to two parallel tangents to the region  $\mathcal{R}$  (see Figure 1). Then  $[\hat{x}_p(\min), p]$  and  $[\hat{x}_p(\max), p]$ ,  $0 < p < 1$ , are two curves in the  $(x, \Phi[(x - \mu)/\sigma])$  plane which define a simultaneous confidence band  $\mathcal{B}$  for all quantiles.

Based on Result 1 in the Appendix, the lower and upper confidence curves for quantiles are the same as the upper and lower confidence curves, respectively, for the cdf  $\Phi$ . That is the band  $\mathcal{B}$  is also the simultaneous confidence band for the cdf. Result 2 in the Appendix shows the band  $\mathcal{B}$  has the same confidence level as the confidence region  $\mathcal{R}$ .

Equation (1) can always be calculated numerically, so the band (2) is obtainable. Cheng and Iles (1983) provide an exact formula for the WALDF method. Suppose  $\gamma$  is the  $1 - \alpha$  quantile of the distribution from the WALDF method and  $\gamma/n < (i_{11}i_{22} - i_{12}^2)/i_{11}$ . Then  $100(1 - \alpha)\%$  two-sided simultaneous confidence curves are

$$C_l(x) = \Phi(\hat{\xi}(x) - h), \quad C_u(x) = \Phi(\hat{\xi}(x) + h) \quad (7)$$

where

$$h = \{\gamma n^{-1} i_{11}^{-1} [1 + (i_{11}\hat{\xi} + i_{12})^2 (i_{11}i_{22} - i_{12}^2)^{-1}]\}^{1/2} \text{ and } \hat{\xi}(x) = (x - \hat{\mu})/\hat{\sigma}.$$

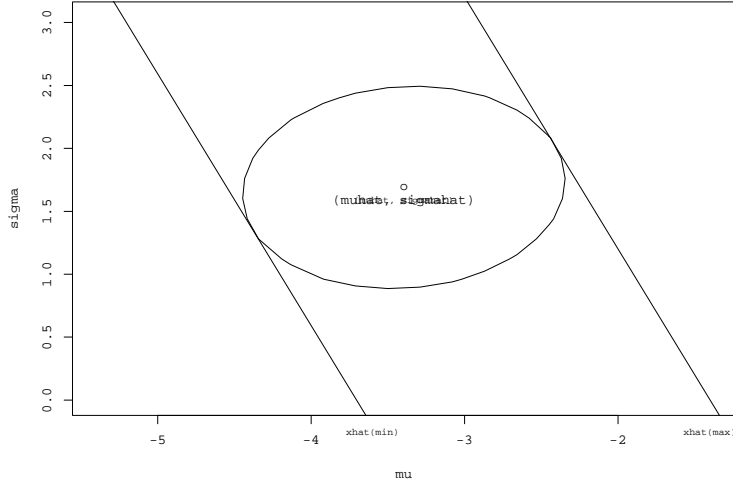


Figure 1: A 95% convex confidence region to be used for two-sided simultaneous confidence band constructed from the BWALDL method with data in Section 5.

The same formula can be applied for the BWALDF method, except that the value of  $\gamma$  is replaced by  $\gamma^*$ , the  $1 - \alpha$  quantile of distribution of  $Q_L^*$ .

Using the arguments similar to those in Cheng and Iles (1983), Escobar and Meeker (2000) develop the following formula for the WALDL method. Their formula also can be used for the BWALDL method. Suppose  $\gamma$  is the  $1 - \alpha$  quantile of the distribution from the WALDL or BWALDL method and  $\gamma/\hat{\sigma}^2 < (\hat{i}_{11}\hat{i}_{22} - \hat{i}_{12}^2)/\hat{i}_{11}$ . Then  $100(1 - \alpha)\%$  two-sided simultaneous confidence curves are

$$C_l(x) = \Phi(\hat{\xi}(x) + h_1 - h_2), \quad C_u(x) = \Phi(\hat{\xi}(x) + h_1 + h_2) \quad (8)$$

where

$$h_1 = \frac{\gamma \times (\hat{i}_{12} + \hat{\xi} \hat{i}_{22})}{\hat{\sigma}^2 - \gamma \times \hat{i}_{22}}$$

$$h_2 = \frac{\sqrt{\gamma}}{\hat{\sigma}^2 - \gamma \times \hat{i}_{22}} \sqrt{\sigma^2 \times (\hat{i}_{11} + 2\hat{\xi} \hat{i}_{12} + \hat{\xi}^2 \hat{i}_{22}) - \gamma \times (\hat{i}_{11}\hat{i}_{22} - \hat{i}_{12}^2)}$$

$$\hat{\xi}(x) = (x - \hat{\mu})/\hat{\sigma}.$$

For the LLR, LLRB, and BLLR methods, the  $100(1 - \alpha)\%$  two-sided simultaneous confidence curves can be obtained numerically by using Equation (1).

### 3.2 One-sided Simultaneous Confidence Bands

The construction of a confidence region for one-sided simultaneous confidence bands is different from the two-sided case. Cheng and Iles (1988) provide an argument for using the

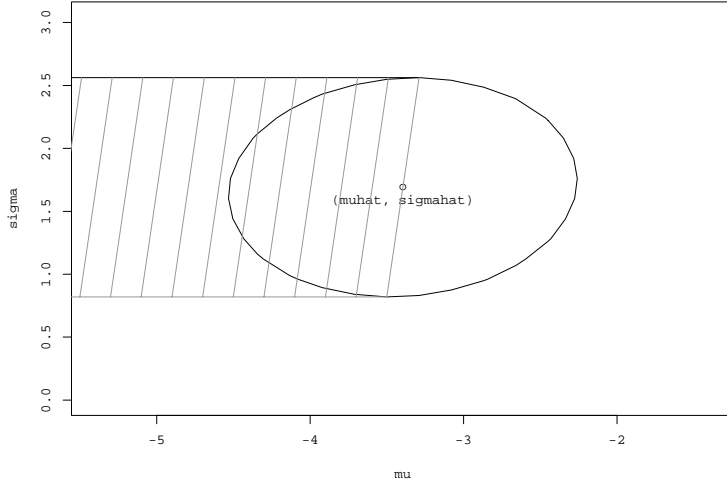


Figure 2: The 97.5% convex confidence region for a one-sided simultaneous confidence band constructed from the BWALDL method with data in Section 5. It is the union of a closed convex region and a left semi-infinite region.

WALDF method. We extend their argument to other methods that can be used to produce convex confidence regions.

To see this, we describe a method for obtaining a region needed to define a lower confidence band of the cdf (the method for an upper confidence band is analogous). As argued in Section 3.1, the upper simultaneous confidence curve for quantiles is the same as the lower simultaneous confidence curve for the cdf. The same argument applies for one-sided simultaneous confidence bands. Below we construct a confidence region for obtaining an upper simultaneous confidence band for quantiles and argue that the confidence level of the band is the same as that of the region.

Suppose we have a convex confidence region  $\mathcal{R}$  with a certain confidence level. For a given  $p$ , let  $\mathcal{R}_p$  denote the half space of  $(\mu, \sigma)$  that satisfies  $\mu + u_p \sigma \leq \hat{x}_p(\max)$ . Let  $\mathcal{R}_l$  denote the intersection of all  $\mathcal{R}_p$ ,  $0 < p < 1$ . Because the tangent lines are on the right boundary of  $\mathcal{R}$ ,  $\mathcal{R}_l$  is the union of region  $\mathcal{R}$  and a left semi-infinite region  $\mathcal{S}_l$ . See Figure 2.

Result 3 (in the Appendix) shows that the confidence level of the lower band  $\mathcal{B}_l$  obtained by using the confidence curve  $C_l$  is the same as that of confidence region  $\mathcal{R}_l$ . That is, the one-sided simultaneous confidence band will be exact if the corresponding convex confidence region  $\mathcal{R}_l$  has the desired confidence level.

We consider  $\mathcal{R}_l$ ,  $\mathcal{R}$ , and  $\mathcal{S}_l$  in their inverted form as being a region  $\mathcal{R}'_l$ ,  $\mathcal{R}'$ , and  $\mathcal{S}'_l$  in the  $(\hat{\mu}, \hat{\sigma})$  plane. For a given confidence coefficient  $1 - \alpha$ , we would like to calculate the corresponding value  $\gamma$  such that

$$\Pr[(\hat{\mu}, \hat{\sigma}) \in \mathcal{R}'_l = \mathcal{R}' \cup \mathcal{S}'_l] = 1 - \alpha. \quad (9)$$

For the WALDF method, Cheng and Iles (1988) describe a way to calculate the critical

value  $\gamma$ . We review and extend their results here. Let

$$\begin{pmatrix} \widehat{\theta}_{F1} \\ \widehat{\theta}_{F2} \end{pmatrix} = \frac{1}{\sigma\sqrt{ni_{11}}} \begin{pmatrix} i_{11} & -i_{12} \\ 0 & d \end{pmatrix} \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\sigma} - \sigma \end{pmatrix},$$

where  $d = (i_{11}i_{22} - i_{12}^2)^{1/2}$ . Note that  $\widehat{\theta}_{F1}$  and  $\widehat{\theta}_{F2}$  are asymptotically independently normal distributed. The region  $\mathcal{R}$  is defined by those  $(\mu, \sigma)$  values that satisfy the inequality  $Q_F = \widehat{\theta}_{F1}^2 + \widehat{\theta}_{F2}^2 \leq \gamma$ . Then spherical symmetry of the independent bivariate normal distribution allows  $\Pr\{(\widehat{\mu}, \widehat{\sigma}) \in \mathcal{R}'_l\}$  to be evaluated as half the sum of  $\Pr\{(\widehat{\mu}, \widehat{\sigma}) \in \mathcal{R}'\}$  and  $\Pr\{(\widehat{\mu}, \widehat{\sigma}) \in \mathcal{S}'_d\}$  where  $\mathcal{S}'_d$  is the doubly infinite band defined by  $|\widehat{\theta}_{F2}| \leq \gamma^{1/2}$ . These probabilities are, respectively,  $\Pr(Z_1 \leq \gamma)$  and  $\Pr(-\gamma^{1/2} \leq Z_2 \leq \gamma^{1/2})$ , where  $Z_1$  is a chi-square random variable with 2 df and  $Z_2$  is a standard normal random variable. Thus the asymptotic value of the confidence coefficient associated with the region  $\mathcal{R}_l$  is given by the formula  $\frac{1}{2}[\Psi(\gamma) + 2\Phi_{\text{nor}}(\gamma^{1/2}) - 1]$ , where  $\Psi$  is the cdf of  $\chi_2^2$  distribution and  $\Phi_{\text{nor}}$  is the cdf of the standard distribution. For the WALDF method, to find the  $\gamma$  for an approximate  $100(1 - \alpha)\%$  confidence region  $\mathcal{R}_l$ , we solve the equation

$$\frac{1}{2}[\Psi(\gamma) + 2\Phi_{\text{nor}}(\gamma^{1/2}) - 1] = 1 - \alpha. \quad (10)$$

For the WALDL method, let

$$\begin{pmatrix} \widehat{\theta}_{L1} \\ \widehat{\theta}_{L2} \end{pmatrix} = \frac{1}{\widehat{\sigma}\sqrt{n\widehat{i}_{11}}} \begin{pmatrix} \widehat{i}_{11} & -\widehat{i}_{12} \\ 0 & \widehat{d} \end{pmatrix} \begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\sigma} - \sigma \end{pmatrix},$$

where  $\widehat{d} = (\widehat{i}_{11}\widehat{i}_{22} - \widehat{i}_{12}^2)^{1/2}$ . By the same argument, we solve the equation (10) to find  $\gamma$  for an approximate  $100(1 - \alpha)\%$  confidence region  $\mathcal{R}_l$ .

For the LLR method, the regions  $\mathcal{R}'$  and  $\mathcal{S}'_d$  are defined by  $W \leq \gamma$  and  $-\gamma^{1/2} \leq R_s \leq \gamma^{1/2}$ , where  $R_s = \text{sign}(\widehat{\sigma} - \sigma)\sqrt{W_1}$ ,  $W_1 = 2[\log L(\widehat{\mu}, \widehat{\sigma}) - \log L(\widehat{\mu}_\sigma, \sigma)]$ ,  $\widehat{\mu}_\sigma$  is the constrained maximum likelihood estimator of  $\mu$  given  $\sigma$ . Because  $W$  and  $R_s$  are, respectively, asymptotically  $\chi_2^2$  and standard normal distributed, we can use equation (10) to find the  $\gamma$  for an approximate  $100(1 - \alpha)\%$  confidence region  $\mathcal{R}_l$ . Note that because both the LLRB and the LLR statistics have the same limiting distribution, the LLRB method has the same  $\gamma$  value as the LLR method.

Based on experience with pointwise confidence intervals (e.g., Jeng and Meeker, 2000), we expect that using bootstrap calibration to obtain  $\gamma$  in the one-sided case will provide a more accurate procedure. Let  $\widehat{\mu}^*$  and  $\widehat{\sigma}^*$ ,  $\mathcal{R}'^*$ ,  $\mathcal{R}^*$ , and  $\mathcal{S}'^*_d$  be the bootstrap versions of  $\widehat{\mu}$  and  $\widehat{\sigma}$ ,  $\mathcal{R}'_l$ ,  $\mathcal{R}'$ , and  $\mathcal{S}'_d$ , respectively. Now  $\mathcal{R}'$  is defined by (5) using any of  $Q_F$ ,  $Q_L$ ,  $W$  or  $W_B$ . For a given confidence coefficient  $1 - \alpha$ , we would like to calculate the corresponding value  $\gamma^*$  such that

$$\Pr[(\widehat{\mu}^*, \widehat{\sigma}^*) \in \mathcal{R}'^*_l = \mathcal{R}'^* \cup \mathcal{S}'^*_d] = 1 - \alpha. \quad (11)$$

Then we use  $\gamma^*$  in place of  $\gamma$  in the WALDF, WALDL, and LLR methods to provide bootstrap confidence regions.

Once the confidence region is constructed, the lower one-sided confidence curve for cdf  $\Phi$  is  $C_l(x) = \min_{\theta \in \mathcal{R}_l} \Phi(x; \theta)$ . Hence the the lower one-sided confidence band is given by equation (3).

Using arguments similar to those in the previous section,  $\hat{\theta}_{F2}$ ,  $\hat{\theta}_{L2}$ , and  $R_s$  are pivotal quantities for complete and Type II censored data. Then the confidence region obtained by bootstrap calibration has exactly the nominal confidence level (except for the Monte Carlo simulation error). Thus the procedure for one-sided simultaneous confidence bands also has the correct coverage probability. For Type I censored data, again we have only approximate results, with the approximation becoming better in large samples.

For calculation of one-sided simultaneous confidence curves, (7) can be used for the WALDF and the BWALDF methods by substituting in the corresponding  $\gamma$  values. Formula (7) can also be used for the WALDL and the BWALDL methods. For the LLR and BLLR method, there is no simple formula but the one-sided simultaneous confidence curves can be calculated numerically from (1).

## 4 Simulation Study

To explore the finite sample performance of these methods, we conducted a simulation experiment using the Weibull distribution and both complete and Type I censored data. Our simulation experiment was designed to study the effects of the following factors:

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

We used 5000 Monte Carlo samples for each  $p_f$  and  $E(r)$  combination. The number of bootstrap replications used was 10000. The levels of the simulation experiment factors used were  $p_f = .01, .1, .5, .9, 1$  and  $E(r) = 3, 5, 7, 10, 15, 20, \text{ and } 30$ . For each Monte Carlo sample we obtained the ML estimates of the location and scale parameters. The confidence regions for the two-sided and one-sided  $100(1 - \alpha)\%$  simultaneous confidence bands were evaluated for  $\alpha = .025$  and  $.05$ . 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  $E(r)$  is small. With  $r = 0$ , ML estimates do not exist. With  $r = 1$ , LR intervals may not exist. Therefore, we calculate the results conditionally on the cases with  $r > 1$ , and report the observed nonzero proportions that resulted in  $r \leq 1$ . See Table 4.

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

From the Figure 3, Figure 4 and other figures obtained by our simulation we have the following results

- Neither the WALDF nor the WALDL method provides an adequate procedure when  $E(r) \leq 30$  for both one-sided and two-sided cases.

Table 1: Number of the cases where  $r = 0$  or 1 in 5000 Monte Carlo simulations of the experiment. The expected numbers rounded to the nearest integer are shown inside parentheses.

		$p_f$			
		.01	.10	.50	.90
$E(r)$	3	984(988)	889(918)	555(546)	132(139)
	5	175(198)	159(168)	54( 53)	1( 2)
	7	34( 35)	24( 27)	6( 4)	0( 0)
	10	2( 2)	2( 1)	0( 0)	0( 0)

- The coverage probability of the LLR method depends both on the sample size and on the expected number of failures. The procedure is adequate when  $E(r) \geq 20$  for both one-sided and two-sided cases.
- The LLRB method is adequate when  $E(r) \geq 5$  for two-sided simultaneous bands and when  $E(r) \geq 30$  for one-sided simultaneous bands. Using a Bartlett correction improves the coverage probability of the procedure for one-sided simultaneous bands only when there is no censoring or slight censoring.
- As expected, the BLLR method is exact for complete data. The coverage probability of the procedure for two-sided simultaneous confidence bands is accurate even in heavily censored cases when  $E(r) = 5$ . But the coverage probability of the procedure for one-sided simultaneous confidence bands is accurate when  $E(r) \geq 15$ .
- As expected, the BWALDF method is exact for complete data. With Type I censoring it is adequate when  $E(r) \geq 10$  for both one-sided and two-sided cases.
- The BWALDL is exact as expected for complete data. With Type I censoring it is adequate when  $E(r) \geq 5$  for both one-sided and two-sided cases.

Overall, the BWALDL method provides the best results. Also there are simple formulas to calculate simultaneous confidence bands for the cdf.

In some situation, the highly discrete behavior of the MLE form Type I censored data can result in less accurate pointwise confidence intervals using bootstrap methods (Jeng and Meeker, 2000). For simultaneous confidence bands for cdf of location-scale model, however, the BWALDL method performs well. Thus, for the location-scale model, the BWALDL method is recommended.

## 5 Simultaneous Confidence Band for POD Curve

Probability of Detection (POD) curves are a commonly used metric for the Non-destructive Evaluation (NDE) capability. We follow the methodology developed by Meeker et al. (1995, 1996, 1997) which is motivated by the need for methods to predict ultrasonic (UT) inspection POD for detecting hard-alpha and other subsurface flaws in titanium using a gated

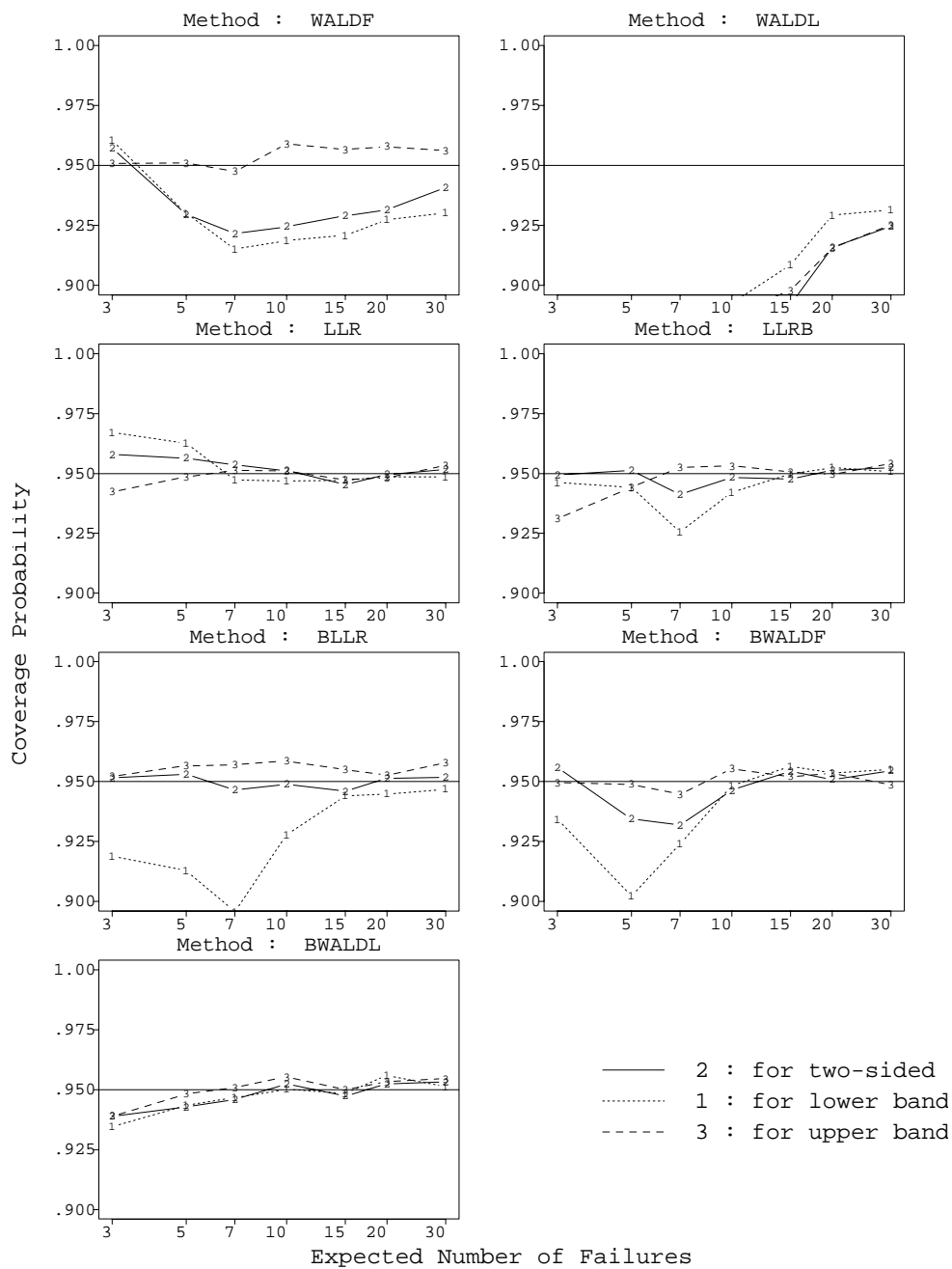


Figure 3: Coverage probability plot of the methods for constructing approximate 95% two-sided and one-sided simultaneous confidence bands in the cases  $p_f = .1$ .

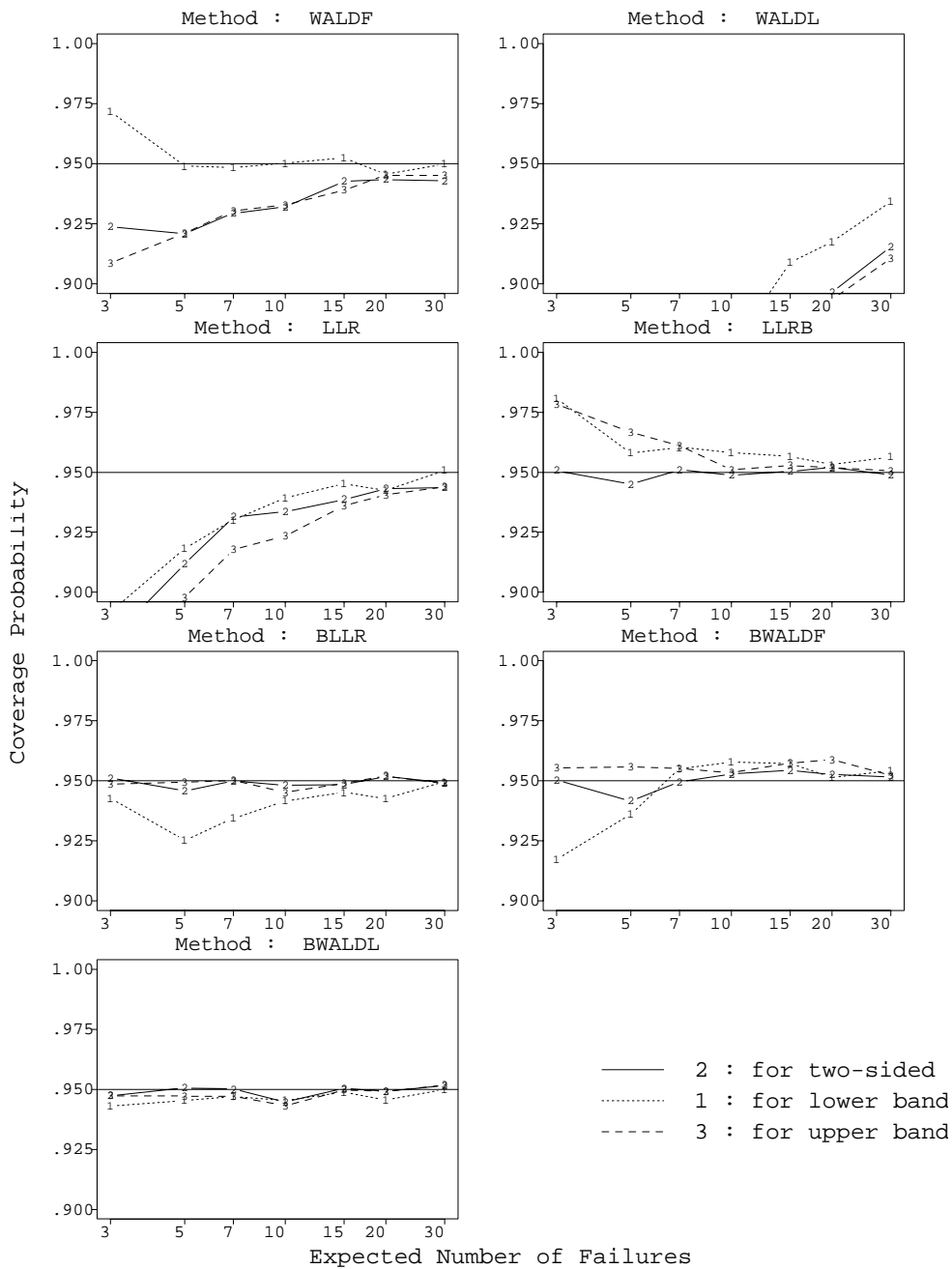


Figure 4: Coverage probability plot of methods for constructing approximate 95% confidence regions for two-sided and one-sided simultaneous confidence bands in the cases  $p_f = .9$ .

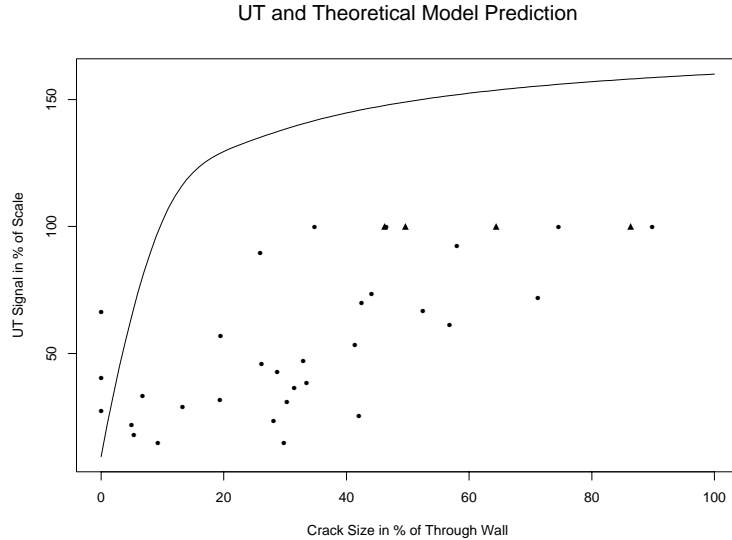


Figure 5: The UT signals and the theoretical model predictions. Censored data is represented by triangles.

peak-to-peak UT detection method. Sarkar et al. (1998) apply a similar methodology to the non-destructive testing using UT inspection and destructive testing of cracks in heat exchanger data.

The data used in Sarkar et al. (1998) came from destructive testing of heat exchanger tubes. The data are denoted by  $\{(a_k, y_k) : k = 1, \dots, n\}$ , where  $y_k$  is the signal amplitude corresponding to the crack size  $a_k$ . In the ultrasonic inspection, the signal will saturate when it exceeds a specific bound. The ultrasonic signals were reported in the scaled format as the percentage of a full-scale signal which is determined by calibrating on a given standard. Signals above 100% are right censored. There are 28 uncensored signals from total 32 observed cracks. The proportion of uncensored signals is .875. Figure 5 shows the observed data and the prediction from a theoretical physical model for an ideal reflecting flaw (a rectangular slot). The signals from an ideal flaw would be much stronger than those from actual cracks. Modeling the deviations between the actual signal strengths and those predicted for the "ideal flaws" provides a useful model for estimating POD for the inspection method.

Let  $\tilde{y}_k$  denote the prediction from the theoretical physical model for ultrasonic NDE signals (UNDE model) for a crack size of  $a_k$ . We define the generalized deviations (using a Box-Cox transformation) as

$$g(y_k, \tilde{y}_k; \lambda) = \begin{cases} \frac{(y_k)^\lambda - 1}{\lambda} - \frac{(\tilde{y}_k)^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(y_k) - \log(\tilde{y}_k), & \lambda = 0. \end{cases} \quad (12)$$

The purpose of using the generalized transformation is to simplify the modeling of variability in the UT signals (specifically to stabilize variance and obtain a simple form for

the distribution). Under the assumption that the distribution of the generalized deviation does not depend on changes in certain inspection parameters (e.g., focus depth or transducer frequency), the model can be used to predict POD for alternative proposed inspection plans.

Based on the experiences with large amounts of experimental UT data, Meeker et al. (1995, 1996) observed that a value of  $\lambda$  in the neighborhood of 0.3 tends to make the distribution of the deviation close to iid  $N(\mu, \sigma^2)$ . For the heat exchanger UT data, Sarkar et al. (1998) also found that  $\lambda = 0.3$  is suitable.

For the heat exchanger UT data, the scaled UT signal amplitude was recorded in the form of single right censoring with the fixed right censoring level  $t_c$ . The generalized deviation results in multiply right-censored values  $x_{ci} = g(t_c, \tilde{y}_i)$ ,  $i = 1, \dots, n$ . We use the method of maximum likelihood to estimate the unknown parameters  $\mu$  and  $\sigma$ . Figure 6 shows a normal probability plot and 95% pointwise confidence intervals for the distribution of the generalized deviations. We see that the normal distribution fits the generalized deviation data well.

Let  $Y$  be the maximum reading in the time gate of an UT A-scan. The threshold  $y_{th}$  was chosen to be the 25% of the full-scale signal. There is a detection when  $Y > y_{th}$ . For this application, the POD is of the primary interest. Under the general model the probability of a detection on any given reading of a crack of size  $a$  is

$$\begin{aligned} POD(a) &= \Pr(Y(a) > y_{th}) = 1 - \Pr[g(Y(a), \tilde{y}(a)) \leq g(y_{th}, \tilde{y}(a))] \\ &= 1 - \Phi_{\text{nor}} \left[ \frac{g(y_{th}, \tilde{y}(a)) - \hat{\mu}_g}{\hat{\sigma}_g} \right] \end{aligned} \quad (13)$$

where  $\Phi_{\text{nor}}$  is the standard normal cumulative distribution function and  $\hat{\mu}_g$  and  $\hat{\sigma}_g$  are estimates from the generalized deviation data.

Sarkar et al. (1998) provide point-wise confidence intervals for a POD curve by using the delta method and a normal approximation. For actual applications of system reliability, one would be interested in the uncertainty of the estimation of the POD curve for a range of crack sizes. The methods developed in this paper provide the needed simultaneous confidence bands.

The following gives the bootstrap procedure used to find the critical value  $\gamma$  for the methods being considered. Let  $U(\mu, \sigma)$  be the particular statistic used for finding the confidence region. This statistic could be  $Q_F$ ,  $Q_L$  or  $W$  that is defined in the Section 3.

1. Simulate one sample  $x_1^*, \dots, x_n^*$  from the normal distribution  $N(\hat{\mu}, \hat{\sigma}^2)$ .
2. Let  $\delta_i^* = 0$  if  $x_i^* \leq x_{ci}$  and  $\delta_i^* = 1$  if  $x_i^* > x_{ci}$ . Set  $z_i^* = \min\{x_i^*, x_{ci}\}$ ,  $i = 1, \dots, N$ . Calculate  $U^*$  from the bootstrap data  $(z_i^*, \delta_i^*)$ ,  $i = 1, \dots, N$ .
3. Repeat steps 1 and 2  $B$  times to calculate the MLEs  $\hat{\mu}_j^*$  and  $\hat{\sigma}_j^*$  of  $\hat{\mu}$  and  $\hat{\sigma}$  and  $U_j^*$  of  $U_j$ ,  $j = 1, \dots, B$ . Arrange the  $U_j$  in ascending order.
4. Use  $U_{B(1-\alpha)}^*$  as the critical value  $\gamma$  for the given confidence coefficient  $1 - \alpha$  in finding two-sided simultaneous confidence bands.
5. Let  $k$  be a positive number less than  $n$  and let  $m = \#\{(\hat{\mu}_j^*, \hat{\sigma}_j^*) : j > k, (\hat{\mu}_j^*, \hat{\sigma}_j^*) \in \mathcal{S}_r'^*\}$  where  $\mathcal{S}_r'^*$  is defined in equation 11. Find  $k$  such that  $k + m = B(1 - \alpha)$ .

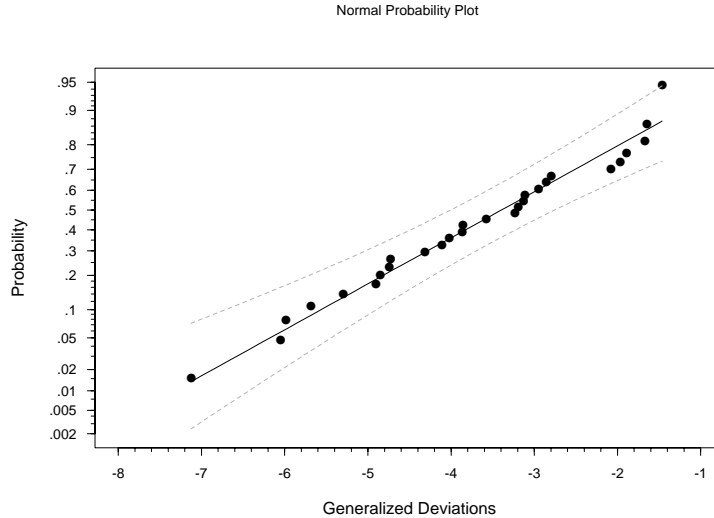


Figure 6: Normal probability plot with  $\lambda = .3$  and 95% pointwise confidence interval.

Use  $U_k^*$  be the critical value  $\gamma$  for the given confidence coefficient  $1 - \alpha$  in finding one-sided simultaneous confidence bands.

We use  $B = 10000$  to calculate the critical value  $\gamma$  for the BWALDF and BWALDL methods and to construct the 95% two-sided and 97.5% one-sided simultaneous confidence bands for the POD curve. Figure 7 compares the two-sided 95% pointwise confidence intervals using the delta method and the simultaneous confidence bands using the BWALDL method. The important differences are clear. The BWALDL band is wider especially when the crack size is smaller than 20% of referenced size. Figure 8 shows that the difference between the BWALDL and the BWALDF methods is not so large. As indicated by the simulation study, because the sample size is 32, the confidence level should be close to the nominal value. Figure 9 compares a set of the 97.5% one-sided lower pointwise intervals based on the normal approximation and lower simultaneous confidence bands based on the BWALDF and BWALDL methods. The pointwise intervals tend to lead to narrower region that would be misleading when interest is over a range of crack sizes.

## 6 Discussion and Future Work

Cheng and Iles (1983, 1988) use Wald statistics with (expected) Fisher information and provide different approaches for finding one-sided and two-sided simultaneous confidence bands when there is no censoring in data. We extend their approach by using Wald statistics with local information and likelihood ratio statistics (with or without Bartlett correction) and compare these to corresponding simulation or bootstrap calibrated versions of the same methods when data are complete or censored. The methods presented in this paper can be used to construct simultaneous confidence bands for general continuous functions of

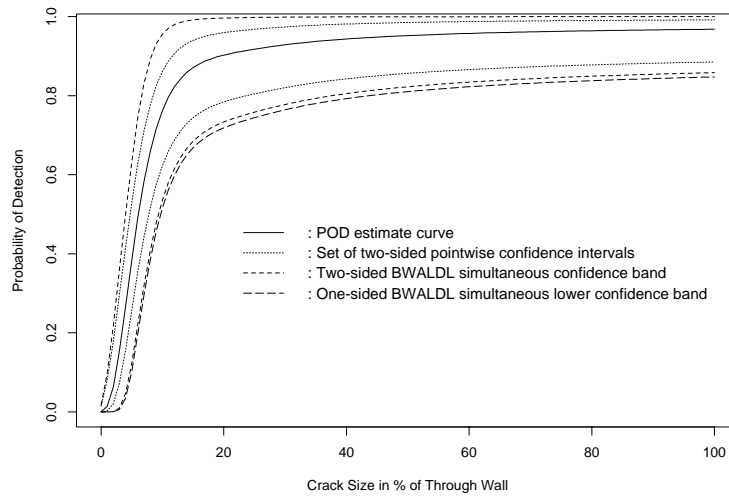


Figure 7: The 95% two-sided pointwise confidence intervals using a normal approximation and the 95% two-sided and 97.5% one-sided lower simultaneous confidence bands using the BWALDL method.

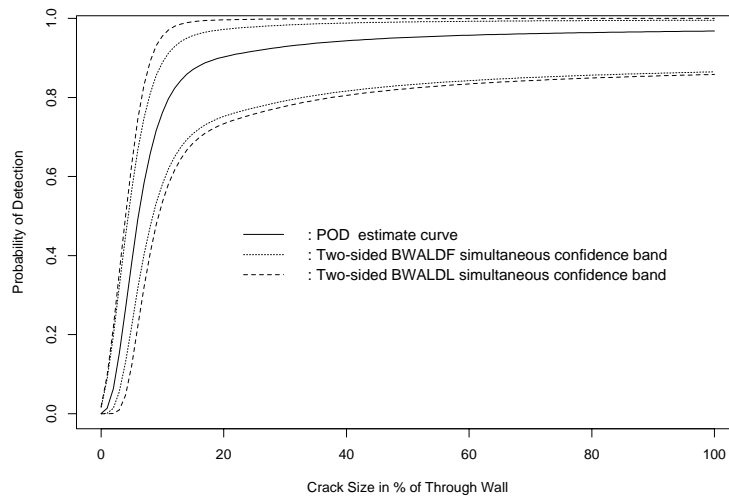


Figure 8: The 95% two-sided simultaneous confidence bands calculated by using the BWALDL and BWALDF methods.

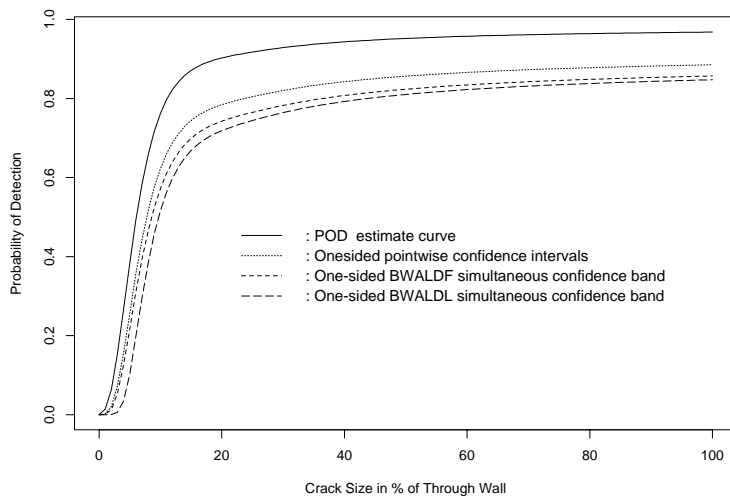


Figure 9: The 97.5% one-sided pointwise lower confidence bounds and simultaneous lower confidence bands calculated by using the BWALDL and BWALDF methods.

unknown model parameters.

We show that for the location-scale model, the accuracy of the procedure for constructing the simultaneous confidence bands is the same as that of the procedure for constructing its corresponding joint confidence region. The BWALDF, BWALDL, and BLLR methods have exact coverage probability when data are complete or Type II censored. When data are Type I censored, only approximate joint confidence regions can be obtained. Our simulation study shows that the BWALDF and BLLR methods provide accurate coverage probabilities when the number of failures reaches 15 for different proportions failing. The BWALDL method produces accurate coverage probabilities when the number of failures reaches 5. The following are some issues for future research:

- In some cases interest centers on inference for a function over some particular range of its arguments. For example, only the lower part of a cdf might be of interest [e.g.,  $\Phi(x; \theta), x < t$ , for some time  $t$ ]. For the cdf of a location-scale distribution, we can construct the corresponding joint confidence regions by following the arguments similar to those in the Sections 3.1 and 3.2 and then use the resulting regions to construct two-sided and one-sided simultaneous bands, respectively. The shape of the joint confidence region will depend on which part of the function is of interest.
- Both the BWALDF and the BWALDL methods provide accurate joint confidence regions for the unknown parameters in the location-scale model. We use these methods to construct correspondingly accurate simultaneous confidence bands for functions of the model parameters. The open question is how well these two methods perform in other models. In particular, it would be useful to know if they are as good as the BLLR method which generally has second order accuracy in coverage probability for

both complete and censored data (Jeng and Meeker, 1999). A general method to construct accurate one-sided simultaneous bands for a function  $g(\cdot; \theta)$  still needs to be explored. The challenge is to determine an appropriate confidence region that can be used to generate a one-sided simultaneous band with nominal coverage probability.

- The approach used to construct simultaneous confidence bands in this paper can be extended to regression problems. Escobar and Meeker (2000) give a formula to calculate the simultaneous confidence band of a regression curve using the WALDF and WALDL methods for the location-scale model. Their formula can also be used for the BWALDF and BWALDL methods. The simultaneous confidence bands using the LLR, LLRB or BLLR methods also can be obtained numerically. We would expect the coverage properties for regression models to be similar to those observed in our simulation study.

## Appendix

### A. Two-sided Simultaneous Confidence Bands

**Result 1.** In a location-scale model, the lower and upper confidence curves for quantiles are the same as the upper and lower confidence curves for the cdf, if those curves are computed from a convex joint confidence region.

**Proof.** We want to show that  $[\hat{x}_p(\min), p]$  and  $[\hat{x}_p(\max), p]$ ,  $0 < p < 1$ , are two curves the same as  $[x, \max_{(\mu, \sigma) \in \mathcal{R}} \Phi((x - \mu)/\sigma)]$  and  $[x, \min_{(\mu, \sigma) \in \mathcal{R}} \Phi((x - \mu)/\sigma)]$ . We only show the lower confidence curve case, the upper case can be obtained analogously. Given  $x$  on real line, there is a  $p$  such that  $\hat{x}_p(\max) = x$ . The lower confidence curve for  $\Phi$  is  $[x, \min_{(\mu, \sigma) \in \mathcal{R}} \Phi((x - \mu)/\sigma)]$ . The claim will be established if we show that  $\min_{(\mu, \sigma) \in \mathcal{R}} \Phi((x - \mu)/\sigma)$  is equal to  $p$ . That is  $(\hat{x}_p(\max), p) = [x, \min_{(\mu, \sigma) \in \mathcal{R}} \Phi((x - \mu)/\sigma)]$ ,  $0 < p < 1$ . Suppose  $\min_{(\mu, \sigma) \in \mathcal{R}} \Phi((x - \mu)/\sigma)$  equals to  $p_0$ . Clearly  $p_0 \leq p$  and there is at least a point  $(\mu_{p_0}, \sigma_{p_0})$  in  $\mathcal{R}$  satisfying equation (6). Suppose that  $p_0$  is smaller than  $p$ . Then it follows that  $-u_{p_0}^{-1}$  is also smaller than  $-u_p^{-1}$ . This implies that the line passing through the point  $(\mu_{p_0}, \sigma_{p_0})$  (inside  $\mathcal{R}$ ) with intercept  $\hat{x}_p(\max)$  is on the right of the tangent line of the region  $\mathcal{R}$  with the same intercept  $\hat{x}_p(\max)$  (see Figure 1 for visual justification). This is impossible. So we have that  $p_0 = p$ .

**Result 2.** In a location-scale model, a two-sided simultaneous confidence band  $\mathcal{B}$  has the same confidence level as its corresponding convex confidence region  $\mathcal{R}$ .

**Proof.** We consider any point  $(\mu_0, \sigma_0)$  which is not in the region  $\mathcal{R}$ . Clearly there is one  $p$  such that the line with slope  $-u_p$  passes the point  $(\mu_0, \sigma_0)$  but does not cross the region  $\mathcal{R}$  (see Figure 2 for visual justification). This implies that the point  $(\mu_0 + u_p \sigma_0, p)$  is not located in the band  $\mathcal{B}$ . So we conclude that no other points outside the region  $\mathcal{R}$  will produce a cdf which lies entirely in the band  $\mathcal{B}$ . The band  $\mathcal{B}$  hence has the same confidence level as the confidence region  $\mathcal{R}$ .

## B. One-sided Simultaneous Confidence Bands

**Result 3.** In a location-scale model, the confidence level of one-sided confidence band of the cdf is the same as the confidence level of its corresponding convex confidence region  $\mathcal{R}_l$ .

**Proof.** We only show the lower confidence band case, the upper case can be obtained analogously. If  $(\mu_0, \sigma_0)$  is not in the region  $\mathcal{R}_l$ , there is at least one  $p_0$  such that the line  $\mu + u_{p_0}\sigma = x_{p_0}$  passing through the point  $(\mu_0, \sigma_0)$  does not cross the region  $\mathcal{R}_l$ . Then the number  $\mu_0 + u_{p_0}\sigma_0$  is bigger than  $\hat{x}_{p_0}(\max)$ . This implies that no other points outside region  $\mathcal{R}_l$  could produce a confidence curve that lies entirely in the band  $\mathcal{B}_l = \{(\hat{x}_p(\max), p) : 0 < p < 1\}$ . Thus the confidence level of  $\mathcal{B}_l$  is the same as that of  $\mathcal{R}_l$ . That is the one-sided simultaneous confidence band will be exact if the corresponding convex confidence region procedure for  $\mathcal{R}_l$  has the nominal confidence level.

## References

- Beran, R. (1988), "Balanced simultaneous confidence sets," *Journal of the American Statistical Association*, **83**, 679-686.
- Cheng, R. C. H. and Iles, T. C. (1983), "Confidence bands for cumulative distribution functions of continuous random variables," *Technometrics*, **25**, 77-86.
- Cheng, R. C. H. and Iles, T. C. (1988), "One-sided confidence bands for cumulative distribution functions" *Technometrics*, **30**, 155-159.
- Cox and Hinkley (1974), *Theoretical Statistics*, London: Chapman & Hall.
- Efron, B., and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall.
- Escobar, L. A. and Meeker, W. Q. (1994), "Fisher information matrix for the extreme value, normal and logistic distributions and censored data", *Applied Statistics*, **43**, 533-540.
- Escobar, L. A. and Meeker, W. Q. (2000), "Simultaneous inference with expected and observed information matrix," Preprint, Department of Statistics, Iowa State University.
- Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.
- Jeng, S. L. and Meeker, W. Q. (1999), "Bootstrap likelihood ratio statistics", Preprint, Department of Statistics, Iowa State University.
- Jeng, S. L. and Meeker, W. Q. (2000), "Comparisons of Approximate Confidence Intervals for Type I Censored Data", *Technometrics*, (will appear in May. 2000)
- Jensen, J. L. (1993), "A historical sketch and some new results on the improved log likelihood ratio statistic," *Scandinavian Journal of Statistics*, **20**, 1-15.
- Kanofsky, P. and Srinivasan, R. (1972), "An approach to the construction of parametric confidence bands on cumulative distribution functions," *Biometrika*, **59**, 623-631.

- Kendall, M. and Stuart A. (1991), *The Kendall's Advanced Theory of Statistics*, Fifth Edition, Vol 2. Oxford University Press.
- Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: John Wiley & Sons.
- Le Cam, L. (1990), "On the standard asymptotic confidence ellipsoids of Wald," *International Statistical Review*, **58**, 129-152.
- Lehmann, E. L. (1986), *Testing Statistical Hypotheses*, New York: John Wiley & Sons.
- Meeker, W. Q., Thompson, R. B., Chiou, C-P, Jeng, S-L, and Tucker, W. T. (1996), "Methodology for estimating nondestructive evaluation capability", *Review of Progress in Quantitative NDE*, **15B**, 1983-1991. D.O. Thompson and D.E. Chimenti, Eds., New York: Plenum Press.
- Meeker, W. Q., Jeng, S-L, Chiou, C-P, and Thompson, R. B. (1997), "Improved methodology for predicting POD of detecting synthetic hard alpha inclusions in Titanium", *Review of Progress in Quantitative Nondestructive Evaluation*, **17B**, 2021-2028. Thompson, D. O. and Chimenti, D. E., Eds., New York: Plenum Press.
- Meeker, W. Q., Jeng, S-L, Chiou, C-P, Thompson, R. B. (1998), "Improved methodology for inspection reliability assessment for detecting synthetic hard alpha inclusions in Titanium", *Review of Progress in Quantitative Nondestructive Evaluation*, Thompson, D. O. and Chimenti, D. E., Eds., **17**, New York: Plenum Press.
- MIL-STD-1823 (1989), "Non-Destructive Evaluation System Reliability Assessment", Standardization Order Desk, Building 4D, 700 Roberts Avenue, Philadelphia, PA 19111-5094.
- Robinson, J. A. (1983), "Bootstrap confidence intervals in location-scale models with progressive censoring," *Technometrics*, **25**, 179-187.
- Rao, C. R. (1973), *Linear Statistical Inference and its Application*, New York: John Wiley & Sons.
- Sarkar, P., Meeker, W. Q., Thompson, R. B., Gray, T. A., and Junker, W. (1998), "Probability of detection modeling for ultrasonic testing," *Review of Progress in Quantitative Nondestructive Evaluation*, Thompson, D. O. and Chimenti, D. E., Eds., **17**, New York: Plenum Press.
- Satten, G. A. (1995), "Upper and Lower Bound Distributions That Give Simultaneous Confidence Intervals for Quantiles", *Journal of American Statistical Association*, **90**, No. 430, 747-752.
- Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons.
- Yeh, A. B. and Singh, K. (1997), "Balanced confidence regions based on Tukey's depth and the bootstrap," *Journal of the Royal Statistical Society, Series B*, **59**, 639-652.