2-11-1999

Statistical Prediction Based on Censored Life Data

Luis A. Escobar
Louisiana State University

William Q. Meeker
Iowa State University, wqmeeker@iastate.edu

Statistical Prediction Based on Censored Life Data

Abstract
This article describes methods for using censored life data to construct prediction bounds or intervals for future outcomes. Both new-sample prediction (e.g., using data from a previous sample to make predictions on the future failure time of a new unit) and within-sample prediction problems (e.g., predicting the number of future failures from a sample, based on early data from that sample) are considered. The general method, based on an assumed parametric distribution, uses simulationbased calibration. This method provides exactly the nominal coverage probability when an exact pivotal-based method exists and a highly accurate large-sample approximation, otherwise. To illustrate new-sample prediction, we show how to construct a prediction interval for a single future observation from a previously sampled population/process (motivated by a customer's request for an interval to contain the life of a purchased product). To illustrate within-sample prediction, we show how to compute a prediction interval for the number of future failures in a specified period beyond the observation period (motivated by a warranty prediction problem). Then we present an example that requires more general methods to deal with complicated censoring arising because units enter service at different points in time (staggered entry).

Keywords
CNDE, Bootstrap, Maximum likelihood, Reliability, Simulation, Warranty

Disciplines
Statistics and Probability

Comments
This preprint has been published in *Technometrics* 41 (1999): 113–124, doi:10.1080/00401706.1999.10485632.

Statistical Prediction Based on Censored Life Data

Luis A. Escobar
Dept. of Experimental Statistics
Louisiana State University
Baton Rouge, LA 70803

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

February 11, 1999
10h 36min

Abstract

This paper describes methods for using censored life data to construct prediction bounds or intervals for future outcomes. Both new-sample prediction (e.g., using data from a previous sample to make predictions on the future failure time of a new unit) and within-sample prediction problems (e.g., predicting the number of future failures from a sample, based on early data from that sample) are considered. The general method, based on an assumed parametric distribution, uses simulation-based calibration. This method provides exactly the nominal coverage probability when an exact pivotal-based method exists and a highly accurate large-sample approximation, otherwise.

To illustrate new-sample prediction we show how to construct a prediction interval for a single future observation from a previously sampled population/process (motivated by a customer's request for an interval to contain the life of a purchased product). To illustrate within-sample prediction, we show how to compute a prediction interval for the number of future failures in a specified period beyond the observation period (motivated by a warranty prediction problem). Then we present an example that requires more general methods to deal with complicated censoring arising because units enter service at different points in time (staggered entry).

Key words: Bootstrap, Maximum Likelihood, Reliability, Simulation, Warranty.

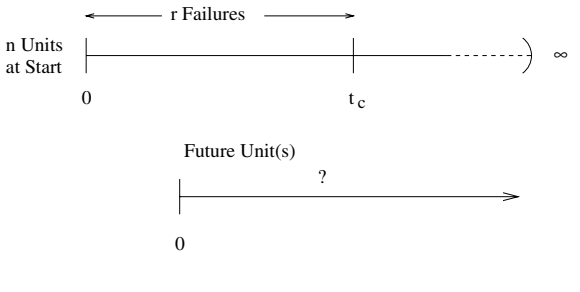

Figure 1: New-sample prediction.

1 Introduction

1.1 Motivation and prediction problems

Practical problems often require the computation of predictions and prediction bounds for future values of random quantities. For example,

- A consumer purchasing a refrigerator would like to have a lower bound for the failure time of the unit to be purchased (with less interest in distribution of the population of units purchased by other consumers).
- Financial managers in manufacturing companies need upper prediction bounds on future warranty costs.
- When planning life tests, engineers may need to predict the number of failures that will occur by the end of the test, or predict the amount of time that it will take for a specified number of units to fail.

Some applications require a two-sided prediction interval $[\underset{\sim}{T}, \quad \widetilde{T}]$ that will, with a specified high degree of confidence, contain the future random variable of interest, say T. In many applications, however, interest is focused on either an upper prediction bound or a lower prediction bound (e.g., the maximum warranty cost is more important than the minimum and the time of the early failures in a product population is more important that the last ones).

Conceptually, it is useful to distinguish between "new-sample" prediction and "within-sample" prediction. For new-sample prediction, data from a past sample is used to make predictions on a future unit or sample of units from the same process or population. For example, based on previous (possibly censored) life test data, one could be interested in predicting the

- Time to failure of a new item.

- Time until k failures in a future sample of m units.
- Number of failures by time t_w in a future sample of m units.

For within-sample prediction, the problem is to predict future events in a sample or process based on early data from that sample or process. For example if n units are followed until t_c and there are r observed failures, $t_{(1)}, \ldots, t_{(r)}$, one could be interested in predicting the

- Time of the next failure, $t_{(r+1)}$.
- Time until k additional failures, $t_{(r+k)}$.
- Number of additional failures in a future interval (t_c, t_w).

1.2 Model

In general to predict a future realization of a random quantity one needs:

- A statistical model to describe the population or process of interest. This model usually consists of a distribution depending on a vector of parameters $\boldsymbol{\theta}$. Nonparametric new-sample prediction is also possible (Chapter 5 of Hahn and Meeker 1991 gives examples and references).
- Information on the values of the parameters $\boldsymbol{\theta}$. This information could come from either a laboratory life test or field data.

We will assume that the failure times follow a continuous distribution with cdf $F(t) = F(t; \boldsymbol{\theta})$ and pdf $f(t) = f(t; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an vector of parameters. Generally, $\boldsymbol{\theta}$ is unknown and will be estimated from available sample data. In such cases we will make the standard assumptions of a) statistical independence of failure times and b) that censoring times are independent of any future failure time that would be observed if a unit were not to be censored (e.g., Section 1.4 of Lawless 1982).

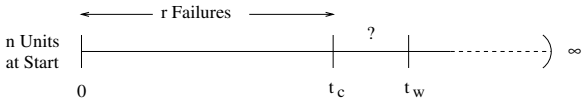

Figure 2: Within-sample prediction.

1.3 Data

The beginning of this paper considers situations in which n units begin operation at time 0 and are observed until a time t_c where the available data are to be analyzed. Failure times are recorded for the r units that fail in the interval $(0, t_c)$. Then the data consist of the r smallest order statistics $t_{(1)} < \cdots < t_{(r)} \leq t_c$ and the information that the other $n - r$ units will have failed after t_c. With time (or Type I) censored data, t_c is prespecified and r is random. With failure (or Type II) censored data, r is prespecified and $t_c = t_{(r)}$ is random. Section 5 shows how to compute prediction bounds for more complicated multiply censored data that are frequently encountered in the analysis of field reliability data.

1.4 Related literature

There is a considerable amount of literature on statistical prediction. Hahn and Nelson (1973), Patel (1989), and Chapter 5 of Hahn and Meeker (1991) provide surveys of methods for statistical prediction for a variety of situations.

Antle and Rademaker (1972) and Nelson and Schmee (1981) provide exact simulation-based prediction interval methods for location-scale (or log-location-scale) distributions and Type II censored data (Type II censoring, however, is rare in practical application). These methods are based on the distribution of pivotal statistics. Engelhardt and Bain (1979) provide a corresponding approximation to the distribution of the required pivotal statistics. Lawless (1973) describes a related conditional method that uses numerical integration. Mee and Kushary (1994) present an alternative simulation-based method that can save important amounts of computer time.

Nagaraja (1995) describes prediction problems for the exponential distribution. He discusses various predictors proposed in the literature and he studies their properties. Nelson (1995) gives a simple procedure for computing prediction limits for the number of failures that will be observed in a future inspection, based on the number of failures in a previous inspection when the units have a Weibull failure-time distribution with a given shape parameter.

Faulkenberry (1973) suggests a method that can be applied when there is a sufficient statistic that can be used as a predictor. Cox (1975) presents a general approximate analytical approach to prediction based on the asymptotic distribution of ML estimators. Atwood (1984) used a similar approach. Efron and Tibshirani (1993, page 390-391) describe an approximate simulation/pivotal-based approach. Beran (1990) gives theoretical results on the properties of prediction statements computed with simulated (bootstrap) samples. Kalbfleisch (1971) describes a likelihood-based method, Thatcher (1964) describes the relationship between Bayesian and frequentist prediction for the bino-

mial distribution, while Geisser (1993) presents a more general overview of the Bayesian approach.

1.5 Overview

This paper explores a simulation-based implementation of the analytical approximate prediction interval procedure suggested by Cox (1975) and studied further by Beran (1990). We illustrate the methods for simple prediction problems and make the connection to the more well-known pivotal-based and approximate pivotal-based methods. Then we illustrate the versatility of the simulation-based method on applications for which neither exact nor approximate pivotal methods exist.

Section 2 describes probability prediction intervals, coverage probability, naive procedures, and other basic ideas pertaining to prediction intervals. Section 3 presents a general approach for calibrating naive statistical prediction intervals. Section 4 shows how to apply the calibration method to a commonly occurring problem of predicting future field failures on the basis of early field failures. Section 5 extends the field prediction problem to situations where units enter the field over a longer period of time (staggered entry). Section 6 contains some concluding remarks and suggestions for further research. The appendix shows the relationship between calibration procedures procedures based on pivotal or pivotal-like statistics as well as some other technical details.

2 Prediction Interval Concepts

2.1 Probability prediction intervals ($\boldsymbol{\theta}$ given)

With a completely specified continuous probability distribution, an exact $100(1-\alpha)\%$ "probability prediction interval" for a future observation from $F(t;\boldsymbol{\theta})$ is (ignoring any data)

$$PI(1-\alpha) = [\underset{\sim}{T}, \quad \widetilde{T}] = [t_{\alpha/2}, \quad t_{1-\alpha/2}] \tag{1}$$

where t_p is the p quantile of $F(t;\boldsymbol{\theta})$. The probability of coverage of the interval in (1) is

$$\Pr[T \in PI(1-\alpha);\boldsymbol{\theta}] = \Pr(\underset{\sim}{T} \leq T \leq \widetilde{T};\boldsymbol{\theta}) = \Pr(t_{\alpha/2} \leq T \leq t_{1-\alpha/2};\boldsymbol{\theta}) = 1-\alpha$$

by the definition of quantiles of continuous distributions.

2.2 Coverage probability for statistical prediction interval procedures ($\boldsymbol{\theta}$ estimated)

Before describing methods for constructing $\boldsymbol{\theta}$-estimated prediction intervals, we first consider methods for evaluating the coverage probability in terms of new-sample prediction of a future failure

time. The ideas also hold, however, for other new-sample prediction problems and for within-sample prediction problems.

In statistical prediction, the objective is to predict the random quantity T based on sample information (denoted by DATA). Generally, with only sample data, there is uncertainty in the distribution parameters. The random DATA leads to a parameter estimate $\widehat{\boldsymbol{\theta}}$ and then to a nominal $100(1-\alpha)\%$ prediction interval $PI(1-\alpha) = [\underset{\sim}{T}, \quad \widetilde{T}]$. Thus $[\underset{\sim}{T}, \quad \widetilde{T}]$ and the future random variable T have a joint distribution that depends on a parameter vector $\boldsymbol{\theta}$.

There are two kinds of coverage probabilities:

- For fixed DATA (and thus fixed $\widehat{\boldsymbol{\theta}}$ and $[\underset{\sim}{T}, \quad \widetilde{T}]$) the conditional coverage probability of a particular interval $[\underset{\sim}{T}, \quad \widetilde{T}]$ is

$$\mathrm{CP}[PI(1-\alpha) \mid \widehat{\boldsymbol{\theta}}; \boldsymbol{\theta}] = \Pr(\underset{\sim}{T} \le T \le \widetilde{T} \mid \widehat{\boldsymbol{\theta}}; \boldsymbol{\theta}) = F(\widetilde{T}; \boldsymbol{\theta}) - F(\underset{\sim}{T}; \boldsymbol{\theta}). \tag{2}$$

  This conditional probability is *unknown* because $F(t; \boldsymbol{\theta})$ depends on the unknown $\boldsymbol{\theta}$.

- From sample to sample, the conditional coverage probability is *random* because $[\underset{\sim}{T}, \quad \widetilde{T}]$ depends on $\widehat{\boldsymbol{\theta}}$. The unconditional coverage probability for the prediction interval *procedure* is

$$\mathrm{CP}[PI(1-\alpha); \boldsymbol{\theta}] = \Pr(\underset{\sim}{T} \le T \le \widetilde{T}; \boldsymbol{\theta}) = \mathrm{E}_{\widehat{\boldsymbol{\theta}}} \left\{ \mathrm{CP}[PI(1-\alpha) \mid \widehat{\boldsymbol{\theta}}; \boldsymbol{\theta}] \right\} \tag{3}$$

  where the expectation is with respect to the random $\widehat{\boldsymbol{\theta}}$. Because it can be computed (at least approximately) and can be controlled, it is this unconditional probability that is generally used to describe a prediction interval procedure.

When $\mathrm{CP}[PI(1-\alpha); \boldsymbol{\theta}] = 1-\alpha$ does not depend on $\boldsymbol{\theta}$, the procedure PI is said to be "exact." When $\mathrm{CP}[PI(1-\alpha); \boldsymbol{\theta}] \neq 1-\alpha$ does not depend on $\boldsymbol{\theta}$, it is generally possible to find a modified procedure PI that is "exact." When $\mathrm{CP}[PI(1-\alpha); \boldsymbol{\theta}]$ depends on the unknown $\boldsymbol{\theta}$, PI is said to be an approximate prediction interval procedure. In such cases it may be possible to modify a specified procedure to find a better approximation.

2.3 Relationship between one-sided prediction bounds and two-sided prediction intervals

Combining a one-sided lower $100(1-\alpha/2)\%$ prediction bound and a one-sided upper $100(1-\alpha/2)\%$ prediction bound gives an equal-tail two-sided $100(1-\alpha)\%$ prediction interval. In particular, if $\Pr(\underset{\sim}{T} \le T < \infty) = 1-\alpha/2$ and $\Pr(0 < T \le \widetilde{T}) = 1-\alpha/2$, then $\Pr(\underset{\sim}{T} \le T \le \widetilde{T}) = 1-\alpha$. It may be possible to find a narrower interval with unequal probabilities in the upper and lower tails,

still summing to α. Use of equal-tail prediction intervals, however, has the important advantage of providing an interval that has endpoints that can be correctly interpreted as one-sided prediction bounds (with the appropriate adjustment in the confidence level). This is important because in most applications the cost of predicting too high is different from the cost of predicting too low and two-sided prediction intervals are often reported even though primary interest is on one side or the other. When computing a two-sided prediction interval, it is often necessary to compute separate lower and upper one-sided prediction bounds and put them together to obtain the prediction interval.

2.4 The naive method for computing a statistical prediction interval

A "naive" prediction interval for continuous T is obtained by substituting the maximum likelihood (ML) estimate for $\boldsymbol{\theta}$ into (1), giving

$$PI(1-\alpha) = [\underset{\sim}{T}, \quad \widetilde{T}] = [\widehat{t}_{\alpha/2}, \quad \widehat{t}_{1-\alpha/2}]$$

where $\widehat{t}_p = t_p(\widehat{\boldsymbol{\theta}})$ is the ML estimate of the p quantile of T. To predict a future independent observation from a log-location-scale distribution (such as the Weibull or lognormal distribution) with cdf $\Pr(T \le t) = \Phi[(\log(t) - \mu)/\sigma]$, a naive prediction interval is

$$\begin{aligned} PI(1-\alpha) &= [\underset{\sim}{T}, \quad \widetilde{T}] = [\widehat{t}_{\alpha/2}, \quad \widehat{t}_{1-\alpha/2}] \\ &= [\exp(\widehat{\mu} + \Phi^{-1}(\alpha/2) \times \widehat{\sigma}), \quad \exp(\widehat{\mu} + \Phi^{-1}(1-\alpha/2) \times \widehat{\sigma})] \end{aligned} \tag{4}$$

where $\Phi(z)$ is the cdf and $\Phi^{-1}(p)$ is the p quantile of the particular standard location-scale distribution. The unconditional coverage probability for this naive procedure is approximately equal to the nominal $1-\alpha$ with large samples sizes. For small to moderate number of units failing, however, the coverage probability may be *far* from $1-\alpha$.

Example 1 Naive prediction interval for predicting the life of a ball bearing (lognormal distribution). Figure 3 is a lognormal probability plot of the first 15 of 23 failures in a bearing life test described in Lawless (1982, page 228) when the data are right-censored at 80 million cycles. Failures occurred at 17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.40, 51.84, 51.96, 54.12, 55.56, 67.80, 68.64, 68.64, and 68.88 million revolutions. The other eight bearings were treated as if they had been censored at 80 million cycles. The lognormal ML estimates are $\widehat{\mu} = 4.160$ and $\widehat{\sigma} = .5451$. From (4), the naive two-sided 90% prediction interval is

$$\begin{aligned} [\underset{\sim}{T}, \quad \widetilde{T}] &= \left[\exp(\widehat{\mu} + \Phi^{-1}_{\text{nor}}(.05) \times \widehat{\sigma}), \quad \exp(\widehat{\mu} + \Phi^{-1}_{\text{nor}}(.95) \times \widehat{\sigma})\right] \\ &= [\exp(4.160 + (-1.645) \times .5451), \quad \exp(4.160 + 1.645 \times .5451)] = [26.1, \quad 157.1]. \end{aligned} \tag{5}$$

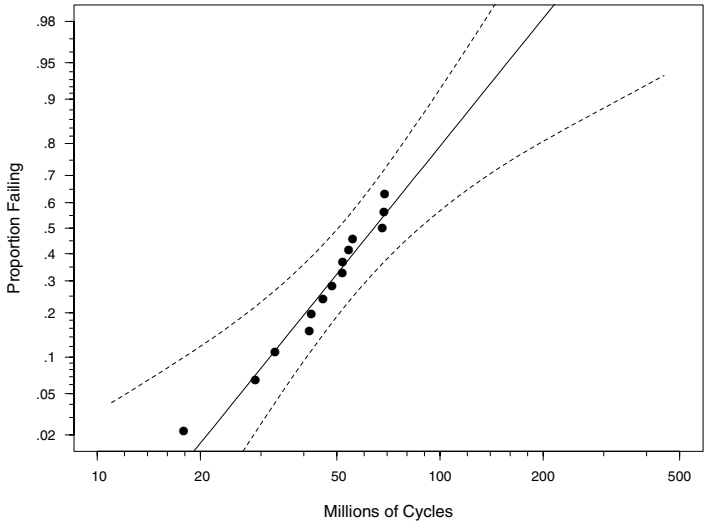

Figure 3: Lognormal probability plot of bearing life test data censored after 80 million cycles (with 15 of 23 units failed) with lognormal ML estimates and pointwise 95% confidence intervals.

Intervals constructed in this manner are generally too narrow and their coverage probability is below the nominal value of $1-\alpha$ because they ignore the uncertainty in $\widehat{\mu}$ and $\widehat{\sigma}$ relative to μ and σ. ∎

3 Calibrating Naive Statistical Prediction Bounds

Cox (1975) suggested a large-sample approximate method, based on maximum likelihood estimates, that can be used to calibrate or correct a naive prediction interval. Atwood (1984) used a similar method. The basic idea of this approach, for a one-sided *lower* confidence bound, is to calibrate the naive one-sided prediction bound by evaluating the function $\mathrm{CP}[PI(1-\alpha_c);\boldsymbol{\theta}]$ at $\widehat{\boldsymbol{\theta}}$ and finding a calibration value $1-\alpha_{cl}$ such that for a one-sided *lower* prediction bound for T

$$\mathrm{CP}[PI(1-\alpha_{cl});\widehat{\boldsymbol{\theta}}] = \Pr\left(\underset{\sim}{T} \le T \le \infty;\widehat{\boldsymbol{\theta}}\right) = \Pr\left(\widehat{t}_{\alpha_{cl}} \le T \le \infty;\widehat{\boldsymbol{\theta}}\right) = 1-\alpha. \tag{6}$$

Here (and hereafter) the notation $\Pr\left(\underset{\sim}{T} \le T \le \infty;\widehat{\boldsymbol{\theta}}\right)$ and $\Pr\left(\widehat{t}_{\alpha_{cl}} \le T \le \infty;\widehat{\boldsymbol{\theta}}\right)$ indicates the corresponding functions $\Pr\left(\underset{\sim}{T} \le T \le \infty;\boldsymbol{\theta}\right)$ and $\Pr\left(\widehat{t}_{\alpha_{cl}} \le T \le \infty;\boldsymbol{\theta}\right)$ evaluated at $\widehat{\boldsymbol{\theta}}$.

Calibration for a one-sided *upper* prediction bound on T (described at the end of Section 3.2) is similar. For a two-sided prediction interval, the calibration is done separately such that the probability is $\alpha/2$ in each tail. Figure 4, to be used in Example 2, provides an illustration of lower and upper "calibration curves."

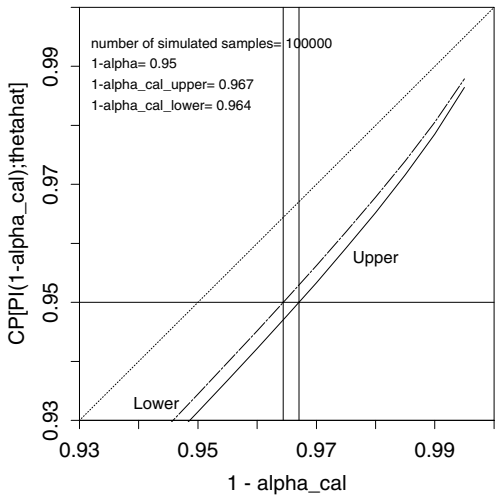

Figure 4: Calibration functions for predicting the failure time of a future bearing based on a lognormal distribution and life test data censored after 80 million cycles.

In problems where $\mathrm{CP}[PI(1-\alpha_c);\boldsymbol{\theta}]$ does not depend on $\boldsymbol{\theta}$, the calibration procedure provides an exact prediction interval. In some simple cases (e.g., prediction based on uncensored samples from exponential and normal distributions), the calibration curve can be obtained analytically in terms of quantiles of standard distributions. Beran (1990) gives examples.

3.1 Approximate calibration of the naive statistical prediction bounds

Cox (1975) suggested an asymptotic analytical approximation for (6). To calibrate a naive lower prediction bound (the method is similar for the upper prediction bound), let $PI(1-\alpha) = [\underset{\sim}{T}, \quad \infty] = [\widehat{t}_\alpha, \quad \infty] = [t_\alpha(\widehat{\boldsymbol{\theta}}), \quad \infty]$. As described in Section 2.2, the conditional coverage probability of $PI(1-\alpha)$ is a function of $1-\alpha$, $\widehat{\boldsymbol{\theta}}$, and $\boldsymbol{\theta}$, say

$$\mathrm{CP}\left[PI(1-\alpha) \mid \widehat{\boldsymbol{\theta}};\boldsymbol{\theta}\right] = \Pr(\underset{\sim}{T} \le T < \infty \mid \widehat{\boldsymbol{\theta}};\boldsymbol{\theta}) = g(\alpha,\widehat{\boldsymbol{\theta}};\boldsymbol{\theta}).$$

Then the unconditional coverage probability of $PI(1-\alpha)$ is $\mathrm{CP}\left[PI(1-\alpha);\boldsymbol{\theta}\right] = \mathrm{E}_{\widehat{\boldsymbol{\theta}}}\left[g(\alpha,\widehat{\boldsymbol{\theta}};\boldsymbol{\theta})\right]$.

Under standard regularity conditions, taking the expectation of a Taylor series expansion of $g(\alpha,\widehat{\boldsymbol{\theta}};\boldsymbol{\theta})$ gives

$$\mathrm{CP}[PI(1-\alpha);\boldsymbol{\theta}] = 1-\alpha + \frac{1}{n}\sum_{i=1}^{k} a_i \left.\frac{\partial g(\alpha,\widehat{\boldsymbol{\theta}};\boldsymbol{\theta})}{\partial\widehat{\theta}_i}\right|_{\boldsymbol{\theta}} + \frac{1}{2n}\sum_{i,j=1}^{k} b_{ij} \left.\frac{\partial^2 g(\alpha,\widehat{\boldsymbol{\theta}};\boldsymbol{\theta})}{\partial\widehat{\theta}_i\,\partial\widehat{\theta}_j}\right|_{\boldsymbol{\theta}} + o\left(\frac{1}{n}\right)$$

where a_i, b_{ij} are elements of the vector $\boldsymbol{a}$ and matrix $\boldsymbol{B}$, respectively, as defined by

$$\begin{aligned} \mathrm{E}_{\widehat{\boldsymbol{\theta}}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})\right] &= \boldsymbol{a}(\boldsymbol{\theta})+o\left(\frac{1}{n}\right) \\ \mathrm{E}_{\widehat{\boldsymbol{\theta}}}\left[(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\right] &= \boldsymbol{B}(\boldsymbol{\theta})+o\left(\frac{1}{n}\right). \end{aligned}$$

Even for seemingly simple problems (e.g., Type I censoring of a sample from a one-parameter exponential distribution), these expectations are extremely difficult to compute. In the few situations where the expressions are tractable (e.g., uncensored samples from exponential and normal distributions), there already exist simple exact prediction procedures based on the distribution of pivotal quantities.

3.2 Calibration by simulation of the sampling/prediction process

Modern computing capabilities make it easy to use Monte Carlo methods to evaluate, numerically, quantities like (6), even for complicated statistical models. Beran (1990) provides asymptotic theory for such prediction calibration. In particular, for a one-sided lower prediction bound, under certain regularity conditions, and with no censoring, Beran shows that the unconditional coverage probability for a once-calibrated prediction procedure $PI(1-\alpha_{cl})$ is

$$\mathrm{CP}[PI(1-\alpha_{cl});\boldsymbol{\theta}] = 1-\alpha+O\left(\frac{1}{n^2}\right).$$

In other words, the dependency of the CP on $\boldsymbol{\theta}$ rapidly diminishes as n increases. The result also holds for one-sided upper prediction bounds and two one-sided bounds used together to form a two-sided prediction interval. Beran (1990) also shows that the order of the asymptotic approximation can be improved by iterating the calibration procedure, but indicated that the performance of the higher-order approximation might not be so good in small samples.

To calibrate with simulation, under the assumed model we can use ML estimates $\widehat{\boldsymbol{\theta}}$ to simulate both the sampling *and* prediction process a large number B (e.g., $B = 50{,}000$ or $B = 100{,}000$) times. Although $B = 2000$ or so is often suggested for simulation-based confidence intervals, larger values of B are generally required for prediction problems due to the added variability of the single future observation.

Calibration of a lower prediction bound. Conceptually, to obtain a calibration curve for a lower prediction bound, like that shown in Figure 4, the function $\mathrm{CP}[PI(1-\alpha_c);\widehat{\boldsymbol{\theta}}]$ in (6), can be evaluated as follows:

1. Choose a particular value of $1-\alpha_c$, say $1-\alpha_0$.

2. Simulate DATA_j^* from the assumed model with parameter values equal to the ML estimates $\widehat{\boldsymbol{\theta}}$ [i.e., from $F(t;\widehat{\boldsymbol{\theta}})$]. Use the sampling procedures and censoring that mimics the original experiment.

3. Compute the simulation ML estimate $\widehat{\boldsymbol{\theta}}_j^*$ from DATA_j^*.

4. Compute the naive $100(1-\alpha_0)\%$ lower prediction bound $\underset{\sim}{T}_j^*$ from the simulated DATA_j^*. Compare $\underset{\sim}{T}_j^*$ with an independent T_j^* simulated from $F(t;\widehat{\boldsymbol{\theta}})$ to see if $T_j^* > \underset{\sim}{T}_j^*$.

5. Repeat steps 2 to 4 for $j = 1, 2, \ldots, B$. The proportion of the B trials having $T_j^* > \underset{\sim}{T}_j^*$ gives the Monte Carlo evaluation of $\text{CP}\,[PI(1-\alpha_0);\boldsymbol{\theta}]$ at $\widehat{\boldsymbol{\theta}}$, which we denote by $\text{CP}^*[PI(1-\alpha_0);\widehat{\boldsymbol{\theta}}]$.

To obtain the calibration curve, repeat steps 2 to 5 for different values of $1-\alpha_0$.

The difference between $\text{CP}\left[PI(1-\alpha_0);\widehat{\boldsymbol{\theta}}\right]$ and $\text{CP}^*\left[PI(1-\alpha_0);\widehat{\boldsymbol{\theta}}\right]$ is due to Monte Carlo error and can be made arbitrarily small by choosing a sufficiently large value of B. To avoid cumbersome notation we will use $\text{CP}\left[PI(1-\alpha_0);\widehat{\boldsymbol{\theta}}\right]$ even when the evaluation is done with simulation.

Operationally, for a log-location-scale distribution where $\boldsymbol{\theta} = (\mu, \sigma)$, the entire $\text{CP}[PI(1-\alpha_c);\widehat{\boldsymbol{\theta}}]$ function in (6) can be evaluated more directly, but equivalently, by using the following procedure:

1. Use simulation to compute B realizations of the pivotal-like statistic $Z_{\log(T^*)} = [\log(T^*) - \widehat{\mu}^*]/\widehat{\sigma}^*$.

2. The empirical distribution of the observed values of the random variable $P = 1 - \Phi[Z_{\log(T^*)}]$ provides a Monte Carlo evaluation of $\text{CP}[PI(1-\alpha_c);\widehat{\boldsymbol{\theta}}]$ in (6). In particular, for a lower prediction bound, $1-\alpha_{cl}$ is the $1-\alpha$ quantile of the distribution of the random variable $P = 1 - \Phi(Z_{\log(T^*)})$.

Calibration of an upper prediction bound. The naive one-sided upper prediction bound for T is calibrated by finding $1-\alpha_{cu}$ such that

$$\text{CP}[PI(1-\alpha_{cu});\widehat{\boldsymbol{\theta}}] = \Pr\left(0 \le T \le \widetilde{T};\widehat{\boldsymbol{\theta}}\right) = \Pr\left(0 \le T \le \widehat{t}_{1-\alpha_{cu}};\widehat{\boldsymbol{\theta}}\right) = 1-\alpha. \tag{7}$$

Then a Monte Carlo evaluation of the entire function $\text{CP}[PI(1-\alpha_c);\widehat{\boldsymbol{\theta}}]$ in (7) can be obtained from the empirical distribution of the observed values of the random variable $P = \Phi[Z_{\log(T^*)}]$. In particular $1-\alpha_{cu}$ is the $1-\alpha$ quantile of the distribution of the random variable $P = \Phi(Z_{\log(T^*)})$.

Appendix Section A.1 provides justification for these procedures and demonstrates the equivalence of the calibration method and the pivotal method for complete and Type II censored data, mentioned in Section 1.4, as well as the corresponding approximate pivotal method that can be used with Type I censoring. For predicting random variables with distributions that are not log-location-scale, the approach is similar, as will be illustrated in Sections 4 and 5.

3.3 Calibration by averaging conditional coverage probabilities

As shown by Mee and Kushary (1994), it can be much more efficient, computationally, to obtain the needed calibration curves for (6) and (7) by simulating conditional coverage probabilities like those in (2) and averaging these to estimate the expectation in (3). The procedure is similar to the one in Section 3.2, replacing steps 4 and 5 with

4. For each simulated sample, compute the *naive* $100(1-\alpha_0)\%$ upper and lower prediction bounds $\underset{\sim}{T}^*$ and $\widetilde{T}^*$, respectively. For a log-location-scale distribution, $\underset{\sim}{T}^* = \exp(\widehat{\mu}^* + \Phi^{-1}(\alpha_0) \times \widehat{\sigma}^*)$ and $\widetilde{T}^* = \exp(\widehat{\mu}^* + \Phi^{-1}(1-\alpha_0) \times \widehat{\sigma}^*)$.

5. A Monte Carlo evaluation of the unconditional coverage probability is obtained from the average of the simulated conditional coverage probabilities $\mathrm{CP}[PI(1-\alpha_0); \widehat{\boldsymbol{\theta}}] = \sum_{j=1}^{B} P_j/B$ where

   (a) For the upper prediction bound calibration $P_j = \Pr(T^* \leq \widetilde{T}^*) = F(\widetilde{T}^*; \widehat{\boldsymbol{\theta}})$. For a log-location-scale distribution, $P_j = \Phi[(\log(\widetilde{T}^*) - \widehat{\mu})/\widehat{\sigma}]$.

   (b) For the lower prediction bound calibration, compute the conditional coverage probability $P_j = \Pr(T^* \geq \underset{\sim}{T}^*) = 1 - F(\underset{\sim}{T}^*; \widehat{\boldsymbol{\theta}})$. For a log-location-scale distribution, $P_j = 1 - \Phi[(\log(\underset{\sim}{T}^*) - \widehat{\mu})/\widehat{\sigma}]$.

To obtain the entire calibration curves, one would need to compute $\mathrm{CP}[PI(1-\alpha_0); \widehat{\boldsymbol{\theta}}]$ for a large number of different values of $1-\alpha_0$ between 0 and 1. Operationally, to compute a one-sided lower prediction bound one needs only to find the appropriate $1-\alpha_{cl}$ value. The $\mathrm{CP}[PI(1-\alpha_c); \widehat{\boldsymbol{\theta}}]$ function is a continuous, increasing function of $1-\alpha_c$, so the appropriate calibration value can be found by using a simple root-finding method.

The procedure for Monte Carlo evaluation of the coverage probability in Section 3.2 utilized the observed proportion of correct prediction bounds. The advantage of the probability-averaging procedure is that it does not require a simulation of the future random variable in the evaluation. Thus the procedure requires fewer Monte Carlo samples to get the same level of accuracy. The method in Section 3.2 might be preferred in situations where a naive prediction interval is easy to compute, but when the conditional probabilities cannot be computed easily (e.g., when the cdf and quantiles of the random variable to be predicted cannot be computed in closed form).

For either evaluation method, it is a simple matter to use standard sampling methods to quantify Monte Carlo error. For example, the standard error of the Monte Carlo evaluation of $\mathrm{CP}[PI(1-$

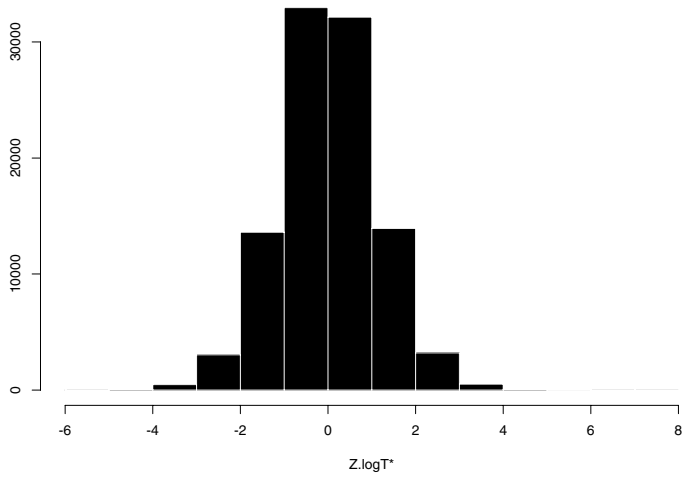

Figure 5: Histogram of 100,000 simulated $Z_{\log(\widehat{T}^*)}$ values, based on the bearing life test data censored after 80 million cycles.

$\alpha_0); \widehat{\boldsymbol{\theta}}]$ for any particular $1-\alpha_0$ is

$$\sqrt{\sum_{j=1}^{B} \frac{(P_j - \mathrm{CP}[PI(1-\alpha_0); \widehat{\boldsymbol{\theta}}])^2}{B(B-1)}}.$$

For the probability-averaging procedure, the variability in the P_j values is related to the variability in $\widehat{\boldsymbol{\theta}}_j^*$ values. The probability-averaging procedure can provide substantial savings in computing time.

Example 2 Calibration of the naive prediction interval for a future lognormal bearing life. Figure 5 is a histogram of the 100,000 simulated values of $Z_{\log(T^*)}$. Figure 6 is a corresponding histogram of the B=100,000 simulated values of $\Phi_{\mathrm{nor}}[Z_{\log(T^*)}]$. The lower and upper $\mathrm{CP}[PI(1-\alpha_c); \widehat{\boldsymbol{\theta}}]$ calibration functions in Figure 4 could have been computed from the empirical cdfs of the simulated $1-\Phi_{\mathrm{nor}}[Z_{\log(T^*)}]$ and $\Phi_{\mathrm{nor}}[Z_{\log(T^*)}]$ values, respectively. Visually, one can imagine integrating the histogram of $\Phi_{\mathrm{nor}}[Z_{\log(T^*)}]$ in Figure 6 and its complement to obtain, respectively, the upper and lower calibration curves shown in Figure 4.

Actually the lower and upper $\mathrm{CP}[PI(1-\alpha_c); \widehat{\boldsymbol{\theta}}]$ calibration functions in Figure 4 were obtained by using the conditional probability averaging method instead, with $B = 100{,}000$. The simulation sample size of $B = 100{,}000$ was chosen to be large enough to assure that the printed calibration values are correct to the number of digits shown. Because B is so large, the differences between the

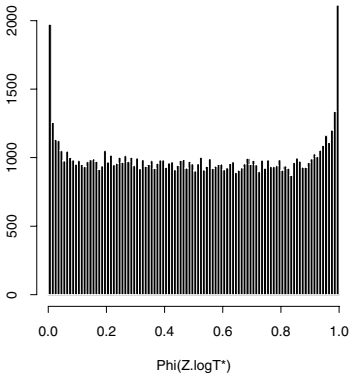

Figure 6: Histogram of 100,000 simulated $\Phi_{\text{nor}}[Z_{\log(\widehat{T}^*)}]$ values, based on the bearing life test data censored after 80 million cycles.

two calibration methods described here were small. With $B = 10{,}000$, the differences were more pronounced, but $B = 10{,}000$ would, for practical purposes, be large enough for the conditional probability averaging method. Using the calibration points in Figure 4, a naive 96.4% lower prediction bound for T provides a calibrated approximate 95% lower prediction bound for T. Also, a naive 96.7% upper prediction bound for T provides a calibrated approximate 95% upper prediction bound for T.

To compute the 90% two-sided prediction interval for T, these two one-sided prediction bounds can be combined. Operationally, substitute $\Phi_{\text{nor}}^{-1}(1-.964) = -1.802$ for $\Phi^{-1}(\alpha/2)$ and $\Phi_{\text{nor}}^{-1}(.967) = 1.837$ for $\Phi^{-1}(1-\alpha/2)$ in (4) giving

$$\begin{aligned}
[\underset{\sim}{T}, \quad \widetilde{T}] &= [\exp(\widehat{\mu} + \Phi_{\text{nor}}^{-1}(1-.964) \times \widehat{\sigma}), \quad \exp(\widehat{\mu} + \Phi_{\text{nor}}^{-1}(.967) \times \widehat{\sigma})] \\
&= [\exp(4.160 + (-1.802) \times .5451), \quad \exp(4.160 + 1.837 \times .5451)] \\
&= [24.0, \quad 174.4] .
\end{aligned}$$

Thus we are 90% confident that the future bearing will fail between 24.0 and 174.4 million cycles of operation.

It is important to note that the upper prediction bound requires some extrapolation given that there were only 15 failures in the sample of 23 of the bearings. This upper bound does not account for possible model error in the unobserved upper tail of the failure-time distribution. ∎

4 Prediction of Future Failures from a Single Group of Units in the Field

Consider the situation where n units are placed into service at approximately one point in time. Failures are reported until t_c, another point in time where the available data are to be analyzed. Suppose that $F(t;\boldsymbol{\theta})$ is used to describe the failure time distribution and that $r > 0$ units have failed in the interval $(0, t_c)$. Thus there are $n - r$ unfailed units at t_c.

A common problem (e.g., in warranty exposure prediction) is the need to predict the number of additional failures K that will be reported between t_c and t_w, where $t_w > t_c$. In addition, it is sometimes necessary to quantify the uncertainty in such a prediction. The upper prediction bound for K is usually of particular interest.

Conditional on the number of failures r, K follows a BINOMIAL$(n - r, \rho)$ distribution where

$$\rho = \frac{\Pr(t_c < T \leq t_w)}{\Pr(T > t_c)} = \frac{F(t_w;\boldsymbol{\theta}) - F(t_c;\boldsymbol{\theta})}{1 - F(t_c;\boldsymbol{\theta})} \tag{8}$$

is the conditional probability of failing in the interval (t_c, t_w), given that a unit survived until t_c. The corresponding binomial cdf is $\Pr(K \leq k) = \text{BINCDF}(k, n - r, \rho)$.

The naive $100(1-\alpha)\%$ upper prediction bound for K is $\widetilde{K}(1-\alpha) = \widehat{K}_{1-\alpha}$. This upper prediction bound is computed as the smallest integer k such that $\text{BINCDF}(k, n-r, \widehat{\rho}) \geq 1-\alpha$. The ML estimate $\widehat{\rho}$ is obtained by evaluating (8) at ML estimate $\widehat{\boldsymbol{\theta}}$. This upper prediction bound can be calibrated by finding $1 - \alpha_{cu}$ such that

$$\text{CP}[PI(1 - \alpha_{cu}); \widehat{\boldsymbol{\theta}}] = \Pr\left[K \leq \widetilde{K}(1 - \alpha_{cu}); \widehat{\boldsymbol{\theta}}\right] = 1 - \alpha. \tag{9}$$

Then the $100(1 - \alpha)\%$ calibrated upper prediction bound would be $\widetilde{K}(1 - \alpha_{cu}) = \widehat{K}_{1-\alpha_{cu}}$.

The naive $100(1 - \alpha)\%$ lower prediction bound for K is $\underset{\sim}{K}(1 - \alpha) = \widehat{K}_{\alpha}$. This naive lower prediction bound is computed as the largest integer k such that $\text{BINCDF}(k, n - r, \widehat{\rho}) < \alpha$. This lower prediction bound can be calibrated by finding $1 - \alpha_{cl}$ such that

$$\text{CP}[PI(1 - \alpha_{cl}); \widehat{\boldsymbol{\theta}}] = \Pr\left[K \geq \underset{\sim}{K}(1 - \alpha_{cl}); \widehat{\boldsymbol{\theta}}\right] = 1 - \alpha \tag{10}$$

and the calibrated lower prediction bound would be $\underset{\sim}{K}(1 - \alpha_{cl}) = \widehat{K}_{\alpha_{cl}}$.

The needed calibration curves for (9) and (10) can be found by averaging conditional coverage probabilities obtained from Monte Carlo simulation by using the following procedure that is similar to the one in Section 3.3.

1. Choose a particular value of $1 - \alpha_c$, say $1 - \alpha_0$.

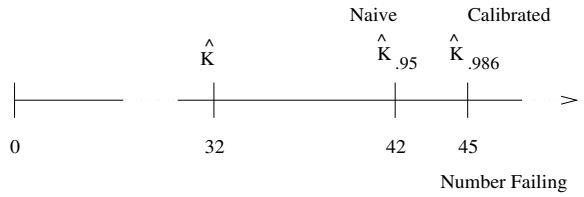

Figure 7: Prediction of the future number failing in the Product-A population.

2. Generate simulated samples of size n, say DATA_j^* for $j = 1, \ldots, B$ from the assumed model with parameter values equal to $\widehat{\boldsymbol{\theta}}$ and the same censoring scheme as in the original sample (leading to the same censoring pattern, except for the variability in $n - r$).

3. The jth simulated sample DATA_j^* provides $n - r_j^*$, $\widehat{\boldsymbol{\theta}}_j^*$, and $\widehat{\rho}_j^*$.

4. Use the cdf $\text{BINCDF}(k; n - r_j^*, \widehat{\rho}_j^*)$ to compute the upper and lower *naive* prediction bounds $\widetilde{K}(1-\alpha_0)_j^*$ and $\underset{\sim}{K}(1-\alpha_0)_j^*$.

5. For the upper prediction bound calibration, compute the conditional coverage probability $P_j = \text{BINCDF}\left[\widetilde{K}(1-\alpha_0)_j^*; n - r_j^*, \widehat{\rho}\right]$. A Monte Carlo evaluation of the unconditional coverage probability is $\text{CP}[PI(1-\alpha_0); \widehat{\boldsymbol{\theta}}] = \sum_{j=1}^{B} P_j / B$.

6. For the lower prediction bound calibration, compute the conditional coverage probability $P_j = 1 - \text{BINCDF}\left[\underset{\sim}{K}(1-\alpha_0)_j^* - 1; n - r_j^*, \widehat{\rho}\right]$. A Monte Carlo evaluation of the unconditional coverage probability is $\text{CP}[PI(1-\alpha_0); \widehat{\boldsymbol{\theta}}] = \sum_{j=1}^{B} P_j / B$.

The justification for this procedure is given in Appendix Section A.2.

Example 3 Prediction interval to contain the number of future Product-A failures. During one month, n =10,000 units of Product-A (the actual name of the product is not being used to protect proprietary information) were put into service. After 48 months, 80 failures had been reported. Management requested a point prediction and an upper prediction bound on the number of the remaining $n - r = 10000 - 80 = 9920$ units that will fail during the next 12 months (i.e., between 48 and 60 months of age). The available data and previous experience suggested a Weibull failure-time distribution and the ML estimates are $\widehat{\alpha} = 1152$ and $\widehat{\beta} = 1.518$. From these,

$$\widehat{\rho} = \frac{\widehat{F}(60) - \widehat{F}(48)}{1 - \widehat{F}(48)} = .003233.$$

Figure 7 shows the point prediction, the naive 95% upper prediction bound, and the calibrated approximate 95% upper prediction bound. The point prediction for the number failing between 48

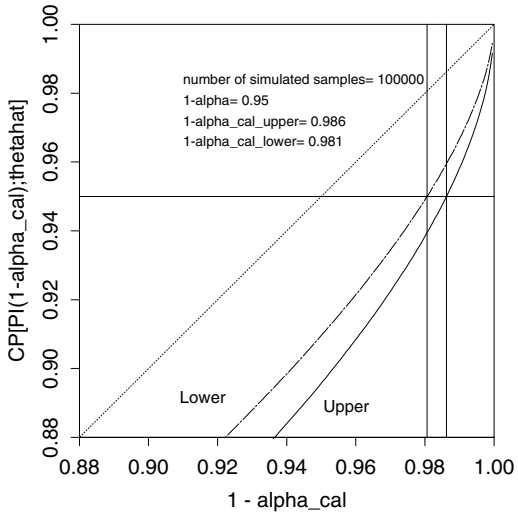

Figure 8: Calibration functions for upper and lower prediction bounds on the number of field failures in the next year for the Product-A population.

and 60 months is $\widehat{K} = (n-r) \times \widehat{\rho} = 9920 \times .003233 = 32.07$. The naive 95% upper prediction bound on K is $\widetilde{K}(.95) = \widehat{K}_{.95} = 42$, the smallest integer k such that BINCDF$(k, 9920, .003233) \geq .95$. The calibration curve shown in Figure 8 gives, for the upper prediction bound, CP$[PI(.986); \widehat{\boldsymbol{\theta}}] = .95$. Thus the calibrated approximate 95% upper prediction bound on K is $\widetilde{K}(.986) = \widehat{K}_{.986} = 45$, the smallest integer k such that BINCDF$(k, 9920, .003233) \geq .986$. The naive 95% lower prediction bound on K is $\underset{\sim}{K}(.95) = \widehat{K}_{.05} = 22$, the largest integer k such that BINCDF$(k, 9920, .003233) < .05$. The calibration curve shown in Figure 8 gives, for the lower prediction bound, CP$[PI(.981); \widehat{\boldsymbol{\theta}}] = .95$. Thus the calibrated approximate 95% lower prediction bound on K is $\underset{\sim}{K}(.981) = \widehat{K}_{.019} = 20$, the largest integer k such that BINCDF$(k, 9920, .003233) < 1 - .981 = .019$. ∎

5 Prediction of Future Failures from Multiple Groups of Units with Staggered Entry into the Field

This section describes a generalization of the prediction problem in Section 4. In many applications the units in the population of interest entered service over a period of time. This is called staggered entry. As in Section 4, the need is to use early field-failure data to construct a prediction interval for the number of future failures in some interval of calendar time, where the amount of previous operating time differs from group to group. This prediction problem is illustrated in Figure 9.

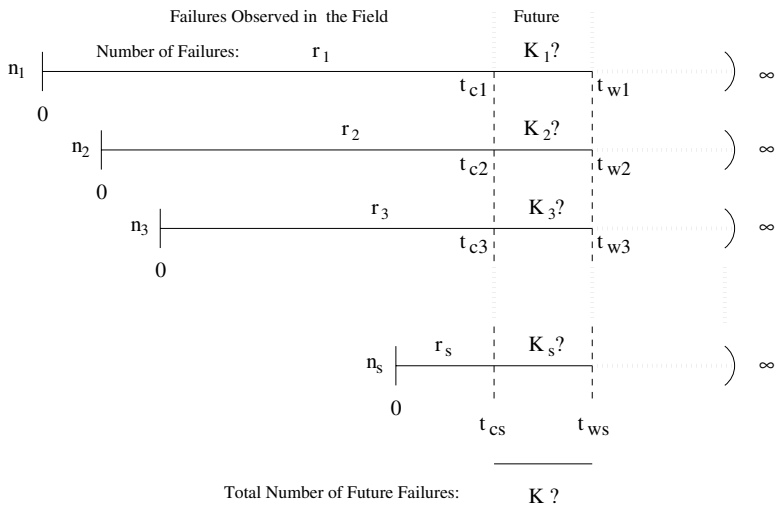

Figure 9: Illustration of staggered entry prediction.

Staggered entry failure-time data are multiply censored because of the differences in operating time. The prediction problem can be viewed as predicting the number of the additional failures across the s groups during a specified period of calendar time. The problem is more complicated than the prediction procedure given in Section 4 because the age of the units, the failure probabilities, and number of units at risk to failure differ from group to group. For group i, n_i units are followed for a period of length t_{ci} and the first r_i failures were observed at times $t_{(i1)} < \cdots < t_{(ir_i)}$, $i = 1, \ldots, s$.

Conditional on $n_i - r_i$, the number of additional failures K_i from group i during interval (t_{ci}, t_{wi}) (where $t_{wi} = t_{ci} + \Delta t$) is distributed BINOMIAL$(n_i - r_i, \rho_i)$ with

$$\rho_i = \frac{\Pr(t_{ci} < T \leq t_{wi})}{\Pr(T > t_{ci})} = \frac{F(t_{wi}; \boldsymbol{\theta}) - F(t_{ci}; \boldsymbol{\theta})}{1 - F(t_{ci}; \boldsymbol{\theta})}. \tag{11}$$

Let $K = \sum_{i=1}^{s} K_i$ be the total number of additional failures over Δt. Conditional on the DATA (and the fixed censoring times) K has a distribution that can be described by the sum of s independent but non-identically distributed binomial random variables with cdf denoted by $\Pr(K \leq k) = \text{SBINCDF}(k; \boldsymbol{n} - \boldsymbol{r}, \boldsymbol{\rho})$ where $\boldsymbol{n} - \boldsymbol{r} = (n_1 - r_1, \ldots, n_s - r_s)$ and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_s)$. Appendix Section A.3 describes methods for evaluating SBINCDF$(k; \boldsymbol{n} - \boldsymbol{r}, \boldsymbol{\rho})$ and the corresponding quantiles of K.

A naive $100(1-\alpha)\%$ upper prediction bound $\widetilde{K}(1-\alpha) = \widehat{K}_{1-\alpha}$ is computed as the smallest integer k such that SBINCDF$(k, \boldsymbol{n} - \boldsymbol{r}^*, \widehat{\boldsymbol{\rho}}^*) \geq 1 - \alpha$. This upper prediction bound can be calibrated by finding $1 - \alpha_{cu}$ such that

$$\text{CP}[PI(1 - \alpha_{cu}); \widehat{\boldsymbol{\theta}}] = \Pr\left[K \leq \widetilde{K}(1 - \alpha_{cu}); \widehat{\boldsymbol{\theta}}\right] = 1 - \alpha.$$

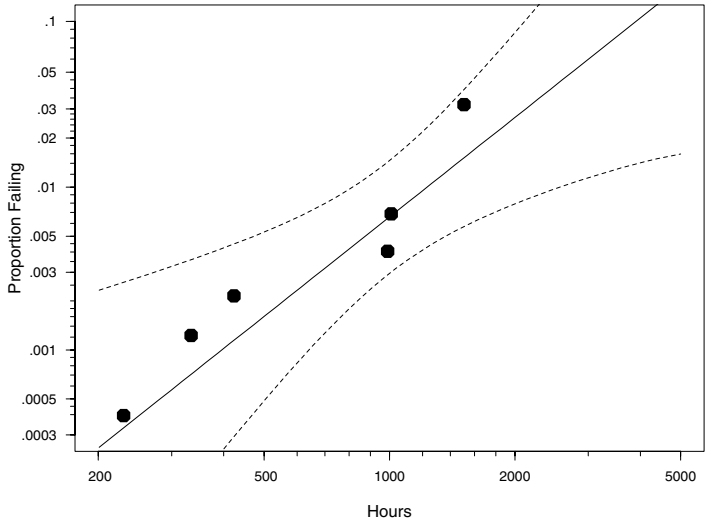

Figure 10: Weibull probability plot of the bearing cage data showing the ML estimate of $F(t)$ along with a set of approximate 95% pointwise confidence intervals for $F(t)$.

A naive $100(1-\alpha)\%$ lower prediction bound $\underset{\sim}{K}(1-\alpha) = \widehat{K}_\alpha$ is computed as the largest integer k such that $\text{SBINCDF}(k, \boldsymbol{n} - \boldsymbol{r}^*, \widehat{\boldsymbol{\rho}}^*) < \alpha$. This lower prediction bound can be calibrated by finding $1 - \alpha_{cl}$ such that

$$\text{CP}[PI(1-\alpha_{cl}); \widehat{\boldsymbol{\theta}}] = \Pr\left[K \geq \underset{\sim}{K}(1-\alpha_{cl}); \widehat{\boldsymbol{\theta}}\right] = 1 - \alpha.$$

To calibrate these one-sided prediction bounds, one can use the same procedure outlined in Section 4, replacing $\text{BINCDF}(k; n-r, \widehat{\rho})$ with $\text{SBINCDF}(k, \boldsymbol{n} - \boldsymbol{r}, \widehat{\boldsymbol{\rho}})$.

Example 4 Prediction interval to contain the number of future bearing cage failures. Abernethy, Breneman, Medlin, and Reinman (1983, pages 43-47) describe the analysis of bearing cage failure data. Groups of bearing cages, installed in a larger system, were introduced into service at different points in time (staggered entry). Failures had occurred at 230, 334, 423, 990, 1009, and 1510 hours of service. There were 1697 other units that had accumulated various amounts of service time without failing. Figure 10 is a Weibull probability plot for the data. Because of an unexpectedly large number of failures in early life, the bearing cage was to be redesigned. It would, however, be some time before the design could be completed, manufacturing started, and the existing units replaced. The analysts wanted to use the initial data to predict the number of additional failures that could be expected from the population of units currently in service, during the next year, assuming that each unit will see $\Delta = 300$ hours of service during the year. Abernethy

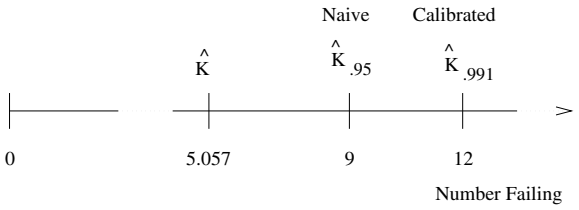

Figure 11: Prediction of the future number failing in the bearing-cage population.

et al. (1983) computed point predictions. We will extend their results to compute a prediction interval to quantify uncertainty.

Table 1 is a future-failure risk analysis. This table gives, for each of the groups of units that had been put into service, the number of units installed, accumulated service times, number of observed failures, estimated conditional probability of failure, and the estimated expected number failing in the 300-hour period. The sum of the estimated expected numbers failing is 5.057, providing a point prediction for the number of failures in the 300-hour period. The Poisson distribution will, in this example, provide a good approximation for the SBIN distribution of K. Figure 11 shows the point prediction, naive upper prediction bound, and the calibrated upper prediction bound for the bearing-cage population. The naive 95% upper prediction bound on K is $\widetilde{K}(.95) = \widehat{K}_{.95} = 9$, the smallest integer k such that $\text{SBINCDF}(k, \boldsymbol{n} - \boldsymbol{r}, \boldsymbol{\rho}) \geq .95$. The upper calibration curve shown in Figure 12 gives, for the upper prediction bound, $\text{CP}[PI(.991); \widehat{\boldsymbol{\theta}}] = .95$. Thus the calibrated 95% upper prediction bound on K is $\widetilde{K}(.991) = \widehat{K}_{.991} = 12$, the smallest integer k such that $\text{SBINCDF}(k, \boldsymbol{n} - \boldsymbol{r}, \boldsymbol{\rho}) \geq .991$. The naive 95% lower prediction bound on K is $\underset{\sim}{K}(.95) = \widehat{K}_{.05} = 1$, the largest integer k such that $\text{SBINCDF}(k, \boldsymbol{n} - \boldsymbol{r}, \boldsymbol{\rho}) < .05$. The lower calibration curve shown in Figure 12 gives $\text{CP}[PI(.959); \widehat{\boldsymbol{\theta}}] = .95$. Thus the calibrated 95% lower prediction bound on K is $\underset{\sim}{K}(.959) = \widehat{K}_{.041} = 1$, the largest integer k such that $\text{SBINCDF}(k, \boldsymbol{n} - \boldsymbol{r}, \boldsymbol{\rho}) < 1 - .959 = .041$. Note that, in this particular case, the naive and the calibrated prediction bounds are the same. ∎

6 Concluding Remarks and Extensions

The methodology described here can be extended in a number of different directions to handle various problems that arise in practice.

- We have illustrated the prediction methods for log-location-scale distributions (such as the Weibull or lognormal distribution). Application to other distributions could follow directly.

Table 1: Bearing cage data and future-failure risk analysis for the next year (300 hours of service per unit).

Group i	Hours in Service	n_i	Failed r_i	At Risk $n_i - r_i$	$\widehat{\rho}_i$	$(n_i - r_i) \times \widehat{\rho}_i$
1	50	288	0	288	.000763	.2196
2	150	148	0	148	.001158	.1714
3	250	125	1	124	.001558	.1932
4	350	112	1	111	.001962	.2178
5	450	107	1	106	.002369	.2511
6	550	99	0	99	.002778	.2750
7	650	110	0	110	.003189	.3508
8	750	114	0	114	.003602	.4106
9	850	119	0	119	.004016	.4779
10	950	128	0	128	.004432	.5673
11	1050	124	2	122	.004848	.5915
12	1150	93	0	93	.005266	.4898
13	1250	47	0	47	.005685	.2672
14	1350	41	0	41	.006105	.2503
15	1450	27	0	27	.006525	.1762
16	1550	12	1	11	.006946	.0764
17	1650	6	0	6	.007368	.0442
18	1750	0	0	0	.007791	0
19	1850	1	0	1	.008214	.0082
20	1950	0	0	0	.008638	0
21	2050	2	0	2	.009062	.0181
Total		1697	6			5.057

Data from Abernethy, Breneman, Medlin, and Reinman (1983, pages 43-47).

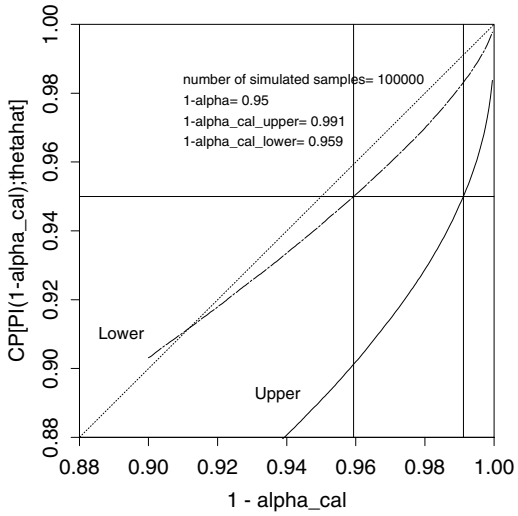

Figure 12: Calibration curve for a prediction interval for the number of bearing cage failures in the next 12 months.

- The calibration methods illustrated here could also be used for problems of simultaneous prediction (e.g., intervals to contain each of several future random variables), and for predicting particular order statistics (in both new-sample and within-sample problems).

- We have illustrated the use of simulation-based prediction methods for two applications that we have encountered in the analysis of product field data. We have seen other variations of these problems. In particular, staggered entry data arise when groups of units are introduced into service every period (commonly, in industry, the number shipped per month is reported). Sometimes there are differences among the underlying failure-time distributions from period to period, resulting from changes in product design. There may, in addition, be strong seasonal effects in the failure process (e.g., in the northern Unites States, there are more automobile battery failures in the winter than in the summer). In some cases there will be end-of-warranty boundaries on the time intervals. These extensions would be straight-forward to handle from a technical point of view, but would require additional bookkeeping and corresponding computer programming beyond that used in the applications presented here.

- In some applications (e.g., failure of outdoor paints and coatings), there may be strong temporal and spatial environmental effects that would have to be considered to obtain accurate prediction bounds.

- Today, the computational price for computing the prediction intervals is relatively small. The difficulty is that each new situation still requires some amount of new programming. It would be useful to have general-purpose software that could easily be adapted to run and use the needed simulations in a time-efficient manner.

Acknowledgments

We have benefited from helpful discussions on this subject with Necip Doganaksoy, Gerry Hahn, Wayne Nelson, and Bill Wunderlin. A referee, an Associate Editor, the Editor, and Shuen-Lin Jeng provided helpful comments on an earlier version of this paper. Computing for the research reported in this paper was done using equipment purchased with funds provided by an NSF SCREMS grant, award DMS 9707740, to Iowa State University.

A Technical Details

This appendix outlines some technical details to justify some of the procedures in the body of the paper.

A.1 Equivalence of the calibration procedure and the pivotal-based (approximate pivotal-based) procedure for prediction intervals from a log-location-scale distribution

This section shows the equivalence of the prediction intervals obtained from the calibration procedure of Section 3.2 and the procedure based on the pivotal (approximate pivotal) quantity $Z_{\log(T)} = [\log(T) - \widehat{\mu}]/\widehat{\sigma}$ when the data are from a log-location-scale distribution.

With Type II (failure) censoring, a life test is run until a specified number of r failures where $1 \leq r \leq n$. For complete or Type II (failure) censored data, $Z_{\log(T)} = [\log(T) - \widehat{\mu}]/\widehat{\sigma}$ is pivotal. That is, the distribution of $Z_{\log(T)}$ depends only on n and r but not on μ and σ. For single time censoring (test run until a specified censoring time t_c), $Z_{\log(T)}$ is only approximately pivotal and quantiles of $Z_{\log(T)}$ depend on $F(t_c; \mu, \sigma)$ (the unknown expected proportion failing by time t_c) and the sample size n. For more information on the pivotal-based method see, for example, Antle and Rademaker (1972), Engelhardt and Bain (1979), and Mee and Kushary (1994). In general, for the

lower prediction bound in a log-location-scale prediction problem

$$\begin{aligned} \mathrm{CP}[PI(1-\alpha_{cl});\mu,\sigma] &= \Pr\left[T \geq \widehat{t}_{\alpha_{cl}};\mu,\sigma\right] \\ &= \Pr\left[\log(T) \geq \widehat{\mu} + \widehat{\sigma}\Phi^{-1}(\alpha_{cl});\mu,\sigma\right] \\ &= \Pr\left[[\log(T)-\widehat{\mu}]/\widehat{\sigma} \geq \Phi^{-1}(\alpha_{cl});\mu,\sigma\right] \\ &= 1-\Pr[Z_{\log(T)} \leq \Phi^{-1}(\alpha_{cl})] = 1-\alpha. \end{aligned} \tag{12}$$

Consequently, $\Pr[Z_{\log(T)} \leq \Phi^{-1}(\alpha_{cl})] = \alpha$, or equivalently, the α quantile of the distribution of $Z_{\log(T)}$ is $\Phi^{-1}(\alpha_{cl})$; that is $z_{\log(T)_{(\alpha)}} = \Phi^{-1}(\alpha_{cl})$. It follows that $1-\alpha_{cl} = 1-\Phi(z_{\log(T)_{(\alpha)}})$ is the $1-\alpha$ quantile of the distribution of $P = \left[1-\Phi(Z_{\log(T)})\right]$.

When $Z_{\log(T)}$ is pivotal, the coverage probability in (12) does not depend on (μ,σ). Thus the the pivotal-based lower prediction bound is $\widehat{\mu}+\widehat{\sigma}\times z_{\log(T)_{(\alpha)}}$ and the calibrated (approximate calibrated) lower prediction bound is $\widehat{\mu}+\widehat{\sigma}\times\Phi^{-1}(\alpha_{cl})$. Noting that $z_{\log(T)_{(\alpha)}} = \Phi^{-1}(\alpha_{cl})$ shows that the two prediction procedures yield the same prediction bound.

When $Z_{\log(T)}$ is not pivotal, (12) is evaluated at the ML estimates $\mu = \widehat{\mu}$ and $\sigma = \widehat{\sigma}$, giving an approximate calibration. This evaluation is expressed as

$$\begin{aligned} \mathrm{CP}^*[PI(1-\alpha_{cl});\widehat{\mu},\widehat{\sigma}] &= \Pr\left[[\log(T^*)-\widehat{\mu}^*]/\widehat{\sigma}^* \geq \Phi^{-1}(\alpha_{cl});\widehat{\mu},\widehat{\sigma}\right] \\ &= 1-\Pr[Z_{\log(T^*)} \leq \Phi^{-1}(\alpha_{cl})] = 1-\alpha. \end{aligned}$$

The approximate pivotal-based lower prediction bound is $\widehat{\mu}+\widehat{\sigma}\times z_{\log(T^*)_{(\alpha)}}$ and the calibrated (approximate calibrated) lower prediction bound is $\widehat{\mu}+\widehat{\sigma}\times\Phi^{-1}(\alpha_{cl})$. Noting that $z_{\log(T^*)_{(\alpha)}} = \Phi^{-1}(\alpha_{cl}))$ shows that the two prediction procedures yield the same prediction bound.

When the quantiles of the distribution of $Z_{\log(T)}$ (or $Z_{\log(T^*)}$) are not available we use simulation to obtain (or approximate) the quantiles of the distribution of these two random variables. For a simulation of size B, the pivotal-based (approximate pivotal-based) procedure uses the α quantile, $z_{\log(T^*)_{(\alpha)}}$, of the empirical distribution of the simulated values $Z_{\log(T_j^*)} = [\log(T_j^*-\widehat{\mu}_j^*)]/\widehat{\sigma}_j^*$, $j = 1,\ldots,B$ to construct the lower prediction bound, $\widehat{\mu}+\widehat{\sigma}\times z_{\log(T^*)_{(\alpha)}}$. Also the empirical distribution of the observed values of $P^* = [1-\Phi(Z_{\log(T^*)})]$ provides an evaluation of $\mathrm{CP}^*[PI(1-\alpha_{cl});\widehat{\mu},\widehat{\sigma}]$. Then, from the results immediately following (12), $1-\alpha_{cl} = [1-\Phi(z_{\log(T^*)_{(\alpha)}})]$ and the calibrated (approximate calibrated) lower prediction bound is $\widehat{\mu}+\widehat{\sigma}\times\Phi^{-1}(\alpha_{cl})$. Again, $z_{\log(T^*)_{(\alpha)}} = \Phi^{-1}(\alpha_{cl})$ showing that the two procedures give identical lower prediction bounds.

When the data are complete or Type II censored, the only differences between the quantiles $z_{\log(T)_{(\alpha)}}$ and $z_{\log(T^*)_{(\alpha)}}$ are due to Monte Carlo error and the coverage of the prediction intervals can be made as close to $1-\alpha$ as desired by taking a large value for the simulation size B.

For the upper prediction bound, the calibration consists of finding $1-\alpha_{cu}$ such that

$$\begin{aligned}\mathrm{CP}[PI(1-\alpha_{cu});\mu,\sigma] &= \Pr\left[T \le \widehat{t}_{1-\alpha_{cu}};\mu,\sigma\right] \\ &= \Pr\left[\log(T) \le \widehat{\mu}+\widehat{\sigma}\Phi^{-1}(1-\alpha_{cu});\mu,\sigma\right] \\ &= \Pr\left[Z_{\log(T)} \le \Phi^{-1}(1-\alpha_{cu});\mu,\sigma\right] = 1-\alpha.\end{aligned}$$

Thus the $1-\alpha$ quantile of the distribution of $Z_{\log(T)}$ is $\Phi^{-1}(1-\alpha_{cu})$. It follows that $1-\alpha_{cu} = \Phi(z_{\log(T)_{(1-\alpha)}})$ is the $1-\alpha$ quantile of the distribution of $P = \Phi(Z_{\log(T)})$. When the quantiles of the distribution of $Z_{\log(T)}$ are not available we again use simulation to approximate the quantiles and the prediction bounds are obtained by replacing $z_{\log(T)_{(1-\alpha)}}$ with $z_{\log(T^*)_{(1-\alpha)}}$ and $1-\alpha_{cu} = \Phi(z_{\log(T)_{(1-\alpha)}})$ with $1-\alpha_{cu} = \Phi(z_{\log(T^*)_{(1-\alpha)}})$.

The pivotal-based (or approximate pivotal-based) upper prediction bound is $\widehat{\mu}+\widehat{\sigma}\times z_{\log(T^*)_{(1-\alpha)}}$ and the (approximate) calibrated upper prediction bound is $\widehat{\mu}+\widehat{\sigma}\times\Phi^{-1}(1-\alpha_{cu})$. Noting that $z_{\log(T^*)_{(1-\alpha)}} = \Phi^{-1}(1-\alpha_{cu})$, this shows that the two prediction procedures yield the same upper prediction bounds.

Note that * denotes a quantity obtained with Monte Carlo but, as indicated in Section 3.2, we have, for the most part, suppressed this notation in the body of the paper.

A.2 Justification of the calibration procedure for prediction bounds for the number of future failures

For a given $1-\alpha_0$, a naive upper prediction bound has the form $\widetilde{K}(1-\alpha_0)$. The unconditional coverage probability evaluated at $\boldsymbol{\theta}$ is

$$\begin{aligned}\mathrm{CP}[PI(1-\alpha_0);\boldsymbol{\theta}] &= \Pr\left[K \le \widetilde{K}(1-\alpha_0);\boldsymbol{\theta}\right] \\ &= \mathrm{E}_{\widehat{\boldsymbol{\theta}}}\left\{\Pr\left[K \le \widetilde{K}(1-\alpha_0) \mid \widehat{\boldsymbol{\theta}};\boldsymbol{\theta}\right]\right\} \\ &= \mathrm{E}_{\widehat{\boldsymbol{\theta}}}\left[\mathrm{BINCDF}(\widetilde{K}(1-\alpha_0);n-r,\rho)\right].\end{aligned}$$

By using simulation, this coverage probability can be evaluated at $\widehat{\boldsymbol{\theta}}$ as follows. For the jth simulated sample of size n, say $\widehat{\boldsymbol{\theta}}^*_j$, the upper prediction bound is $\widetilde{K}(1-\alpha_0)^*_j$ and the conditional coverage probability of the upper prediction bound is $P^*_j = \mathrm{BINCDF}(\widetilde{K}(1-\alpha_0)^*_j;n-r^*_j,\widehat{\rho})$. Using the B simulated samples, the unconditional coverage probability $\mathrm{CP}[PI(1-\alpha_0);\widehat{\boldsymbol{\theta}}]$ is approximated by $\mathrm{CP}^*[PI(1-\alpha_0);\widehat{\boldsymbol{\theta}}] = \sum_{j=1}^{B} P^*_j/B$. The calibration problem consists of finding $1-\alpha_{cu}$ such that $\mathrm{CP}^*[PI(1-\alpha_{cu});\widehat{\boldsymbol{\theta}}] = 1-\alpha$.

Similarly, the unconditional coverage probability corresponding to a lower prediction bound

$\underset{\sim}{K}(1-\alpha_0)$ is

$$\mathrm{CP}[PI(1-\alpha_0);\boldsymbol{\theta}] \quad = \quad 1 - \mathrm{E}_{\widehat{\boldsymbol{\theta}}}\left\{ \mathrm{BINCDF}\left[\underset{\sim}{K}(1-\alpha_0) - 1; n-r, \rho \right] \right\}.$$

Using evaluation at $\widehat{\boldsymbol{\theta}}$, the calibration problem problem is finding $1-\alpha_{cl}$ such that $\mathrm{CP}^*[PI(1-\alpha_{cl});\widehat{\boldsymbol{\theta}}] = \sum_{j=1}^{B} P_j^*/B = 1-\alpha$, where $P_j^* = 1 - \mathrm{BINCDF}\left[\underset{\sim}{K}(1-\alpha_{cl})_j - 1; n-r_j^*, \widehat{\rho} \right]$.

Note that in either case, for fixed $\widehat{\boldsymbol{\theta}}$, $\mathrm{CP}^*[PI(1-\alpha_c);\widehat{\boldsymbol{\theta}}]$ is a continuous function of $1-\alpha_c$.

A.3 Evaluation of the distribution of the sum of s independent non-identically distributed binomial random variables

This appendix describes some methods for evaluating $\mathrm{SBINCDF}(k;\boldsymbol{m},\boldsymbol{\rho})$, the cdf of $K = \sum_{i=1}^{s} K_i$, the sum of s independent non-identically distributed binomial random variables. Here $\boldsymbol{m} = (m_1,\ldots,m_s)$ and $\boldsymbol{\rho} = (\rho_1,\ldots,\rho_s)$ are the number of trials and probabilities for the s different binomial distributions. In general there is not a simple closed form expression for $\mathrm{SBINCDF}(k;\boldsymbol{m},\boldsymbol{\rho})$. If the number of groups is small (e.g., $s \leq 3$), then one can write a convolution formula for the cdf. The complexity of the expression, however, grows exponentially with s and for values of s larger than 3 or 4 it will be useful to consider alternative methods of computation.

If $\rho_i << m_i, i = 1,s$ then one can use the Poisson approximation $K_i \stackrel{\cdot}{\sim} \mathrm{POISSON}(\lambda_i)$, where $\lambda_i = m_i\rho_i$. Thus $\mathrm{SBINCDF}(k;\boldsymbol{m},\boldsymbol{\rho}) \approx \mathrm{POISSON}(\lambda)$, where $\lambda = \sum_{i=1}^{s}\lambda_i$ can be used. If $5 < \rho_i m_i < m_i - 5, i = 1,s$ then by the central limit theorem $\mathrm{SBINCDF}(k;\boldsymbol{m},\boldsymbol{\rho}) \approx \mathrm{NOR}(\mu,\sigma)$, should provide an adequate approximation where $\mu = \sum_{i=1}^{s} m_i\rho_i$ and $\sigma = [\sum_{i=1}^{s} m_i\rho_i(1-\rho_i)]^{1/2}$. In general, $\mathrm{SBINCDF}(k;\boldsymbol{m},\boldsymbol{\rho})$ can be evaluated to any degree of accuracy using Monte Carlo simulation. To approximate $\mathrm{SBINCDF}(k;\boldsymbol{m},\boldsymbol{\rho})$ with Monte Carlo simulation,

1. Generate v_{ju} from $\mathrm{BINOMIAL}(m_u,\rho_u)$, for $u = 1,\ldots,s$.
2. Compute $v_j = \sum_{u=1}^{s} v_{ju}$.
3. Repeat steps 1 and 2 for $j = 1,\ldots,M$ where M should be chosen to be large enough to keep the Monte Carlo error small.
4. The empirical cdf of $v_1,\ldots,v_M$ approximates the cdf $\mathrm{SBINCDF}(k;\boldsymbol{m},\boldsymbol{\rho})$.

Quantiles of the SBINCDF distribution can also be obtained from this empirical distribution.

The Monte Carlo approach will require more computer time than the simple approximations, but less than direct evaluation when s is large.

References

Abernethy, R. B., Breneman, J. E., Medlin, C. H., and Reinman, G. L. (1983), *Weibull Analysis Handbook*, Air Force Wright Aeronautical Laboratories Technical Report AFWAL-TR-83-2079. Available from the National Technical Information Service, Washington, DC.

Antle, C. E. and Rademaker, F. (1972), An upper confidence limit on the maximum of m future observations from a Type I extreme value distribution, *Biometrika* **59,** 475-477.

Atwood, C. L. (1984), Approximate tolerance intervals, based on maximum likelihood estimates, *Journal of the American Statistical Association* **79,** 459-465.

Beran R. (1990), "Calibrating prediction regions," *Journal of the American Statistical Association,* Vol. 85, 715–723.

Cox, D. R. (1975), "Prediction intervals and empirical Bayes confidence intervals," in *Perspectives in Probability and Statistics*, J. Gani, Editor, London: Academic Press.

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

Engelhardt, M. and Bain, L. J. (1979). Prediction limits and two-sample problems with complete or censored Weibull data, *Technometrics* **21,** 233-237.

Faulkenberry, G. D. (1973), A method of obtaining prediction intervals, *Journal of the American Statistical Association* **68**, 433-435.

Geisser, S. (1993), *Predictive Inference: An Introduction*, New York : Chapman and Hall

Hahn, G. J., and Meeker, W.Q. (1991), *Statistical Intervals: A Guide for Practitioners*, New York: John Wiley & Sons, Inc.

Hahn, G. J., and Nelson, W. (1973), A survey of prediction intervals and their applications, *Journal of Quality Technology* **5**, 178-188.

Kalbfleisch, J. D. (1971), Likelihood methods of prediction, in *Foundations of Statistical Inference; Proceedings of the Symposium on the Foundations of Statistical Inference*, edited by Godambe, V.P. and Sprott, D.A. , Toronto: Holt, Rinehart, and Winston of Canada.

Lawless, J. F. (1973), On estimation of safe life when the underlying distribution is Weibull, *Technometrics* **15,** 857-865.

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

Mee, R. and Kushary, D. (1994), Prediction limits for the Weibull distribution utilizing simulation, *Computational Statistics and Data Analysis*, **17** 327-336.

Nagaraja, H. N. (1995), Prediction Problems, Chapter 9 in *The Exponential Distribution: Theory, Methods & Applications*, N. Balakrishnan and A. P. Basu, Editors. Gordon and Breach Publishers.

Nelson, W. (1995), Weibull prediction of a future number of failures, Paper presented at the 1995 Joint Statistical Meetings, Toronto, August 1995.

Nelson, W. and Schmee, J. (1981), Prediction limits for the last failure of a (log) normal sample from early failures, *IEEE Transactions on Reliability* **R-30**, 461-463.

Patel, J. K. (1989), Prediction intervals–a review, *Communications in Statistics–Theory and Methods* **18**, 2393-2465.

Thatcher, A. R. (1964), Relationships between Bayesian and confidence limits for prediction (with discussion), *Journal of the Royal Statistical Society* **B26**, 176-210.