8-24-2018

Applications of the Fractional-Random-Weight Bootstrap

Chris Gotwalt
SAS Institute, Inc.

Li Xu
Virginia Tech

Yili Hong
Virginia Tech

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

Applications of the Fractional-Random-Weight Bootstrap

Abstract

The bootstrap, based on resampling, has, for several decades, been a widely used method for computing confidence intervals for applications where no exact method is available and when sample sizes are not large enough to be able to rely on easy-to-compute large-sample approximate methods, such a Wald (normal-approximation) confidence intervals. Simulation based bootstrap intervals have been proven useful in that their actual coverage probabilities are close to the nominal confidence level in small samples. Small samples analytical approximations such as the Wald method, however, tend to have coverage probabilities that greatly exceed the nominal confidence level. There are, however, many applications where the resampling bootstrap method cannot be used. These include situations where the data are heavily censored, logistic regression when the success response is a rare event or where there is insufficient mixing of successes and failures across the explanatory variable(s), and designed experiments where the number of parameters is close to the number of observations. The thing that these three situations have in common is that there may be a substantial proportion of the resamples where is not possible to estimate all of the parameters in the model. This paper reviews the fractional-random-weight bootstrap method and demonstrates how it can be used to avoid these problems and construct confidence intervals. For the examples, it is seen that the fractional-random-weight bootstrap method is easy to use and has advantages over the resampling method in many challenging applications.

Disciplines

Statistics and Probability

Applications of the Fractional-Random-Weight Bootstrap

Chris Gotwalt
JMP Division, SAS
Research Triangle, NC 12345

Li Xu and Yili Hong
Department of Statistics
Virginia Tech
Blacksburg, VA 24061

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

August 27, 2018

Abstract

The bootstrap, based on resampling, has, for several decades, been a widely used method for computing confidence intervals for applications where no exact method is available and when sample sizes are not large enough to be able to rely on easy-to-compute large-sample approximate methods, such a Wald (normal-approximation) confidence intervals. Simulation based bootstrap intervals have been proven useful in that their actual coverage probabilities are close to the nominal confidence level in small samples. Small samples analytical approximations such as the Wald method, however, tend to have coverage probabilities that greatly exceed the nominal confidence level. There are, however, many applications where the resampling bootstrap method cannot be used. These include situations where the data are heavily censored, logistic regression when the success response is a rare event or where there is insufficient mixing of successes and failures across the explanatory variable(s), and designed experiments where the number of parameters is close to the number of observations. The thing that these three situations have in common is that there may be a substantial proportion of the resamples where is not possible to estimate all of the parameters in the model. This paper reviews the fractional-random-weight bootstrap method and demonstrates how it can be used to avoid these problems and construct confidence intervals. For the examples, it is seen that the fractional-random-weight bootstrap method is easy to use and has advantages over the resampling method in many challenging applications.

Key Words: Censored data, Confidence interval, Design of experiments, Prediction interval, Resampling, Variable selection.

1

1   Introduction

1.1   Bootstrap Background

The bootstrap is a popular statistical tool that is used, primarily, to obtain improved inferences such as approximate confidence intervals and approximate prediction intervals that have coverage probabilities that are close to the nominal confidence level. Bootstrapping is a set of procedures for sampling from the distribution of an estimator that most commonly employs various data generation and augmentation procedures to create new data sets from which new individual samples of the estimator are computed. These samples of the estimators can then be used for many purposes, including approximate confidence and prediction intervals that have more desirable inferential properties than their more commonly used deterministic counterparts. With modern computing technology (hardware and software) bootstrap methods are easy to implement and use. Bootstrap procedures can be applied even in situations where classical theory offers little or no guidance about how to compute trustworthy confidence interval. Generally, there are only minimal regularity conditions (such as a finite variance and a certain degree of smoothness) that are needed to make bootstrap methods work well. Technical details of bootstrap methods can be found in classical references such as Hall (1992), Efron and Tibshirani (1993), Shao and Tu (1995), and Davison and Hinkley (1997).

There are many different types of bootstrap procedures which can be broadly partitioned into nonparametric and parametric. Nonparametric bootstrap procedures require no assumptions about the shape of the underlying data-generating probability distribution. Most bootstrap procedures do assume sampling from a continuous or approximately continuous distribution. Bootstrapping is most commonly done via a Monte Carlo simulation. The most common approach is to generate a sequence of new data sets using resampling. In resampling, each new data set is generated by sampling the rows of the original data with replacement. This approach for generating the bootstrap samples is nonparametric because no assumption about the shape of the underlying distribution is required.

Bootstrap samples can also be generated by assuming a particular parametric distribution and simulating from that distribution. In applications where censoring or truncation is involved, censoring and truncation must be done in a manner that mimics what was done in the original data-generating process. For example, if censoring is random, then a model for the censoring variable needs to be used in the parametric simulation. Often details about how data were censored are not known or is complicated and in such situations, the nonparametric resampling method is much easier to implement. After each bootstrap data set is generated, the statistical procedure (e.g., model fitting and computation of point estimates and in some cases standard errors) is applied to the bootstrap dataset and results are stored. This bootstrap-

sample generation/estimation procedure is repeated a number of times (e.g., 2,500 times) and then the saved results are processed to make inferences (e.g., construct confidence intervals). There are many different ways to use bootstrap samples to compute a confidence interval (e.g., simple percentile, bias-corrected percentile, bias-corrected and accelerated, percentile-t intervals).

1.2   The Idea of Data Weights

In many data analysis applications, it is convenient to put weights (also known as frequencies or counts) on observations. For example, binary data such as 0010001000100010001 are usually replaced with counts of the number of zeros and ones. Weights are frequently used in life test data. For example, the data typically consist of the failure times (which all have weight 1 unless there are ties—often caused by failures being recorded at discrete inspection times) plus the number of units that survived a 1,000-hour test. Weights are also used in data compression where data are binned and the weights indicate the number of observations in each bin (e.g., as displayed in a histogram). Also, when observations have known non-constant variances, it is appropriate to use weights that are inversely proportional to the variance of each observation. The resampling bootstrap method of generating bootstrap samples can also be viewed as data with random *integer* weights. That is, each observation has a weight indicating the number of times it was drawn in the resample. We will give an explicit example of this in the next section.

Many statistical estimation methods allow the specification of weights or frequencies. For example, consider a data set $\boldsymbol{x}_n = (x_1, x_2, \ldots, x_n)$ with corresponding weights $\boldsymbol{w}_n = (w_1, w_2, \ldots, w_n)$. Then estimates of the mean (μ) and variance (σ^2) can be computed from

$$\widehat{\mu} = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i x_i, \quad \text{and} \quad \widehat{\sigma}^2 = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i (x_i - \widehat{\mu})^2.$$

There are similar equations for more general weighted least squares for linear regression models. More generally, suppose we have a dataset $\boldsymbol{x}_n = (x_1, x_2, \ldots, x_n)$ with corresponding weights $\boldsymbol{w}_n = (w_1, w_2, \ldots, w_n)$, where each x_i may contain information such as a response, explanatory variables, and censoring or truncation indicators for observation i. Then the weighted likelihood is

$$L(\boldsymbol{\theta}; \boldsymbol{x}_n, \boldsymbol{w}_n) = \mathcal{C} \prod_{i=1}^n [L_i(\boldsymbol{\theta}; x_i)]^{w_i},$$

where $\boldsymbol{\theta}$ is a general notation for the unknown parameters, $\mathcal{C}$ is a constant that does not related to $\boldsymbol{\theta}$, and $L_i(\boldsymbol{\theta}; x_i)$ is the likelihood contribution from observation i.

In general, we can see that the data weight idea is common in statistical methods and it provides an easy way for computational implementations. The data weight idea also pro-

vides an alternative way to understand bootstrap methods. The objective of this paper is to review the random *fractional-weight* bootstrap method and demonstrate how to apply it to applications in which the resampling (integer weight) bootstrap methods tend to not work well. These applications include heavily censored data, logistic regression when the success response is a rare event or where there is insufficient mixing of successes and failures across the explanatory variable(s), and designed experiments where the number of parameters is close to the number of observations.

1.3   Literature Review

Much has been written about the bootstrap methods since their introduction in the late 1970s. For example, the textbooks by Efron and Tibshirani (1993), and Davison and Hinkley (1997) describe bootstrap theory and methods. The books by Hall (1992) and Shao and Tu (1995) focus on the theory behind bootstrap methods. Another notable reference, aimed at teaching bootstrap methods, is Hesterberg (2015).

As there are only a handful of articles devoted to it, in spite of its usefulness, the FRW bootstrap sampling method appears to be quite under-appreciated at this point. We believe the FRW bootstrap method could serve a much larger role in the toolkit of the applied statistician. Barbe and Bertail (1995) provide a highly technical presentation of the asymptotic theory of various random-weight methods for generating bootstrap estimates. They show how to choose the distribution of the random weights by using Edgeworth expansions. Chatterjee and Bose (2005) present a generalized bootstrap for which the traditional resampling and various weighted likelihood and other weighted estimating equation methods are special cases. Chiang et al. (2005) apply the FRW bootstrap methods to a recurrent events application with informative censoring in a semi-parametric model. Hong, Meeker, and McCalley (2009) apply FRW bootstrap methods to a prediction interval application involving complicated censoring and truncation. Xu, Hong, and Meeker (2015) use the FRW bootstrap in a prediction application to assess the risk of future failures.

1.4   Overview

The remainder of this paper is organized as follows. Section 2 introduces the concept of integer and fractional-weight bootstrap methods. Section 3 gives some theoretical properties of the FRW bootstrap method. Section 4 provides applications of the FRW bootstrap in confidence intervals, which involves heavily censored field-failure data, current-status censored data, and data that are sensitive in estimating the shape parameter of the generalized gamma distribution. Section 5 illustrates the application of the FRW bootstrap to compute prediction

intervals using an application to predict failure of power transformers. Section 6 describes an example where the FRW bootstrap is used to find an appropriate model to describe the results of a designed experiment. Section 7 provides some concluding remarks and areas for further research. Technical details are given in the appendix.

2 Integer and Fractional-Weight Bootstrap

2.1 Integer-weight Bootstrap

Under the idea of data weights, the commonly-used resampling bootstrap procedure is equivalent to choosing the weights from a multinomial distribution with uniform probability $1/n$ for each of the original observations in the sample, where n is the number of observations. That is, the weights $(w_1, \ldots, w_n)'$ follows a uniform multinomial distribution that has a mean 1 and a variance $(n-1)/n$.

As an illustration, the first column of Table 1 gives data tree volume for 15 loblolly pine trees in units of cubic meters. The data are a subsample of data analyzed in Chapter 13 of Meeker, Hahn, and Escobar (2017). The other three columns give the results of resampling with replacement from the sample of size 15, indicating the number of times that each tree was selected for each of the three resamples. As described in Section 1, in an actual application of the bootstrap the resampling would be done B times, usually on the order of thousands. Then a weighted estimation method could be applied to each bootstrap resample to obtain the B bootstrap estimates.

2.2 Fractional-weight Bootstrap Samples

Extending the idea of integer weights, the FRW is introduced with continuous weights. In this case, the weight vector $(w_1, \ldots, w_n)'$ is generated from a from a uniform Dirichlet distribution, multiplied by n. The probability density function (pdf) of the Dirichlet distribution of order n with parameters $\alpha_1, \ldots, \alpha_n$ is given by

$$f(w_1, \ldots, w_n; \alpha_1, \ldots, \alpha_n) = \frac{1}{B(\alpha_1, \ldots, \alpha_n)} \prod_{i=1}^{n} w_i^{\alpha_i - 1}, \qquad \sum_{i=1}^{n} w_i = 1, \qquad w_i \geq 0, \quad (1)$$

and $B(\alpha_1, \ldots, \alpha_n)$ is the normalizing factor. The uniform Dirichlet distribution is a special case when $\alpha_i = 1, i = 1, \ldots, n$. The continuous weights, like the integer multinomial resampling weights, will sum to n, have expectation 1, but a variance $(n-1)/(n+1)$. As an illustration, the last three columns of Table 1 shows the random fractional weights drawn from a uniform Dirichlet distribution, multiplied by n.

Table 1: Three integer-weight (on the left) and fractional-weight (on the right) bootstrap samples.

Tree Volume	Uniform Multinomial Distribution Integer Weights			Uniform Dirichlet Distribution Continuous Weights		
	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
0.149	1	1	1	0.203	0.485	1.451
0.086	2	0	0	0.065	1.328	2.062
0.149	3	0	0	0.629	1.737	0.676
0.194	0	0	1	0.505	0.953	0.590
0.044	1	1	0	0.735	1.510	0.580
0.104	1	1	1	2.543	0.320	2.512
0.156	0	2	1	2.650	0.714	1.320
0.122	1	0	1	0.690	2.072	0.650
0.117	0	3	2	1.095	0.017	0.901
0.079	3	0	2	2.075	1.344	0.792
0.179	0	0	1	0.020	2.368	0.061
0.307	0	7	0	1.947	0.116	1.917
0.049	0	0	1	1.433	0.633	0.982
0.165	1	0	2	0.131	1.137	0.212
0.043	2	0	2	0.279	0.265	0.294
Sum	15	15	15	15	15	15

The FRW bootstrap was first suggested by Rubin (1981). He called it the Bayesian bootstrap because, as shown in the paper, estimates computed from the FRW bootstrap samples are draws from a posterior distribution under a particular relatively diffuse prior distribution. Newton and Raftery (1994) generalized Rubin's ideas and introduced the weighted likelihood bootstrap, which is easy to implement. Newton and Raftery (1994) also show that the weighted likelihood bootstrap is first order accurate.

Even though these random-weight bootstrap methods were developed within a nonparametric Bayesian framework, they also apply to non-Bayesian and parametric inference problems, as will be illustrated in the examples in this paper. There are statistically valid alternative methods to generate the random fractional weights. In particular, Jin, Ying, and Wei (2001) show that FRW bootstrap estimators have good properties if positive independent and identically distributed (iid) weights are generated from a continuous distribution that has a mean and standard deviation equal to one (e.g., an exponential distribution with mean one). Then the sum of the weights is a random variable with expectation n.

The FRW or Bayesian bootstrap is also known as: the random-weight bootstrap, the weighted likelihood bootstrap, the weighted bootstrap, and the perturbation bootstrap. Operationally, the FRW bootstrap samples are used in the same way as the resampling bootstrap samples. Like resampling, the method is nonparametric. There are, however, important advantages of using the FRW bootstrap in certain common applications. The advantages arise because all of the original observations remain in all of the bootstrap samples. In situations where dropping certain observations from a data set will cause estimation problems, the resampling bootstrap approach will often give poor results or fail altogether. Generally, when using the FRW bootstrap, because all of the original observations remain in the sample, estimation difficulties do not arise.

3   Theoretical Results

In this section, we present some theoretical results, which provide the basis for the statistical inference for the FRW bootstrap. For likelihood based inference, it can be shown that the fractional weights generated from the uniform Dirichlet distribution are equivalent to generating standardized random weights from an exponential distribution with mean one. Let $Z_i, i = 1, \ldots, n$ be independent and identically distributed exponential distribution with mean one. Then the random vector

$$\left(\frac{Z_1}{\sum_{i=1}^n Z_i}, \ldots, \frac{Z_i}{\sum_{i=1}^n Z_i} \cdots, \frac{Z_n}{\sum_{i=1}^n Z_i} \right)'$$

(2)

has a uniform Dirichlet distribution.

For convenience of the development of the theoretical properties, we use the weights from the exponential distribution with mean one. Let $X_1, X_2, \ldots, X_n$ be n random iid observations and $\boldsymbol{X}_n$ a general notation for the collection of the n random observations. The loglikelihood function can be written as

$$\bar{l}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} l_i(\boldsymbol{\theta}; X_i),$$

where $l_i(\boldsymbol{\theta}; X_i)$ is the contribution for observation i and $\boldsymbol{\theta}$ is a general notation for the vector of unknown parameters. The maximum likelihood (ML) estimate $\widehat{\boldsymbol{\theta}}$ is the solution to the first derivative $\bar{l}'(\boldsymbol{\theta}) = \partial\bar{l}(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$. The random weighted loglikelihood is

$$\bar{l}^*(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} Z_i l_i(\boldsymbol{\theta}; X_i).$$

Note that the term $\sum_{i=1}^{n} Z_i$ in (2) is ignored because it will not affect the solution. The FRW version of the ML estimate $\widehat{\boldsymbol{\theta}}^*$ is the solution to $\bar{l}^{*\prime}(\boldsymbol{\theta}) = \partial\bar{l}^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = 0$. The proofs of the following three results are given in the appendix.

Result 1 *The FRW ML estimator $\widehat{\boldsymbol{\theta}}^*$ is consistent for $\boldsymbol{\theta}$ if $\widehat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}$. That is if $\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}$ then $\widehat{\boldsymbol{\theta}}^* \to \boldsymbol{\theta}$, as $n \to \infty$.*

Note that the ML estimator $\widehat{\boldsymbol{\theta}}$ is consistent and asymptotically unbiased under some mild conditions (e.g., pages 309-310 of Cox and Hinkley 1974). **Result 1** shows that the FRW bootstrap estimator is also consistent, and thus it is also asymptotically unbiased, which is a desirable property as sometimes bootstrap methods can generate estimates that are biased.

The asymptotic normality is related to the distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})|\boldsymbol{X}_n$ (i.e., given the random sample).

Result 2 *The distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})$ conditional on the random sample $\boldsymbol{X}_n$ goes to $N[0, I(\boldsymbol{\theta})^{-1}]$ as $n \to \infty$. That is $\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})|\boldsymbol{X}_n \to N[0, I(\boldsymbol{\theta})^{-1}]$, as $n \to \infty$. Here $I(\boldsymbol{\theta})$ is the Fisher information matrix for $\boldsymbol{\theta}$ based on $\boldsymbol{X}_n$.*

Note that the ML estimator $\widehat{\boldsymbol{\theta}}$ asymptotically has a $N[\boldsymbol{\theta}, I(\boldsymbol{\theta})^{-1}]$ distribution, under some mild conditions. **Result 2** shows that the distributions of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}^*$ are the same when n goes to ∞. Thus, one can use the distribution of a function of $\widehat{\boldsymbol{\theta}}^*$ to mimic the distribution of the corresponding function of $\widehat{\boldsymbol{\theta}}$.

Under mild conditions, the ML estimates exist for the FRW samples for the log-location-scale family of distributions with right censoring. Specifically, consider data $(t_i, \delta_i), i = 1, \ldots, n$ where t_i is the time to event, δ_i is the censoring indicator, and n is the number of data points. The parameters are denoted by $\boldsymbol{\theta} = (\mu, \sigma)'$ where μ is the location parameter and σ is the scale parameter. The loglikelihood can be written as $l(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta})$, where $l_i(\boldsymbol{\theta})$ is the log

likelihood contribution from observation i. The weighted loglikelihood is $l^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} w_i l_i(\boldsymbol{\theta})$ (here, the normalized weights w_i are used for convenience).

Result 3 *For data with right censoring that are generated from commonly used log-location-scale family of distributions (e.g., lognormal and the Weibull), the minimum condition for the ML estimate to exist for $l^*(\boldsymbol{\theta})$, is either (i) two distinct failure times t_1 and t_2, or (ii) one failure time t_1 and a right-censored observation t_2 with $t_2 > t_1$.*

Because of the continuous weights (i.e., all w_i's are positive), a failure will always make a contribution to the likelihood in the FRW samples. **Result 3** indicates that the requirement for the existence of the ML estimate is mild.

4 Applications to Confidence Intervals

In this section, we use three examples to illustrate the applications of FRW in the construction of confidence intervals.

4.1 Bearing Cage Field Failure Data

4.1.1 Background

There were 1703 aircraft engines that had been put into service over time, as shown in the event plot in Figure 1(a). There had been 6 failures and there were 1697 right-censored observations. These data were originally given in Abernethy et al. (1983) and were re-analyzed in Chapter 8 of Meeker and Escobar (1998).

4.1.2 Weibull Analysis

The Weibull distribution with cdf

$$F(t; \eta, \beta) = \Pr(T \le t) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \ t > 0,$$

will be used to describe these data, where T is the time to failure, $\eta = \exp(\mu)$ is the scale parameter, and $\beta = 1/\sigma$ is the shape parameter, while μ and σ are location and scale parameters respectively for $\log(T)$. Figure 1(b) is a Weibull probability plot of the field-failure data. Table 2 summarizes the numerical results of the estimation. For this example, we will focus on the estimation of the Weibull shape parameter β. The ML estimate is 2.035. The upper endpoint of the Wald 95% confidence interval is 5.67. Because of the small number of failures, the Wald confidence interval is not trustworthy. The likelihood upper endpoint
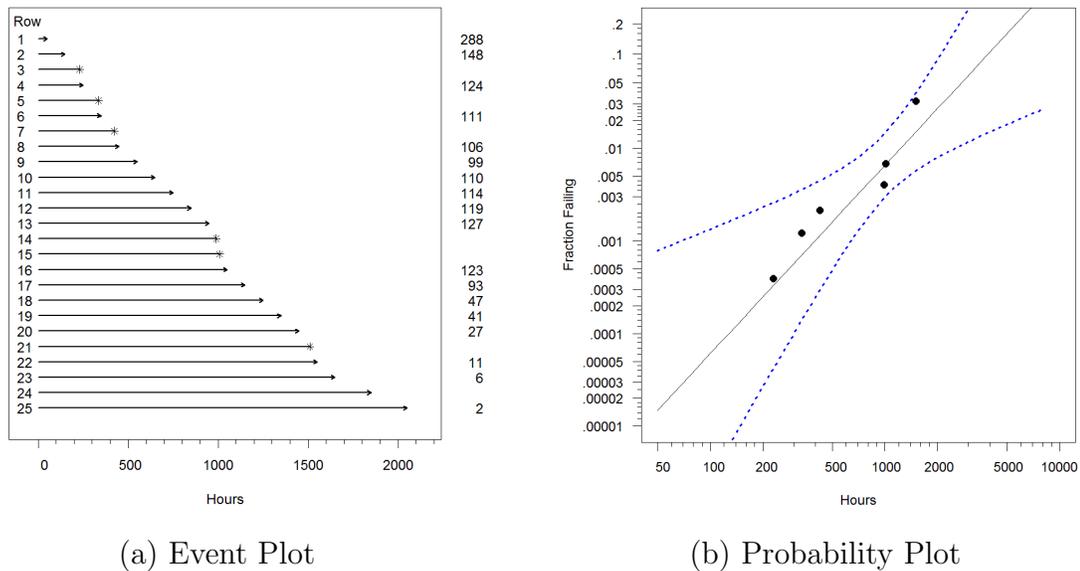
(a) Event Plot

(b) Probability Plot

Figure 1: Event plot and Weibull probability plot for the bearing cage field-failure data.

Table 2: ML estimates for the Weibull analysis of the ball bearing life test data.

| Parameter | Estimate | Std Error | 95% CI | |
			Lower	Upper
η	11792.178	9848.1267	2294.6744	60599.215
β	2.035	0.6657	1.2403	5.670

is 3.58. Another alternative for computing trustworthy confidence intervals is the bootstrap. Care is needed, however, when using the resampling bootstrap method with heavy censoring. If the expected number failing is too small there could be bootstrap samples with only 0 or 1 failures, possibly causing the ML algorithm to fail. As described in **Result 3** of Section 3, there is a unique maximum of the likelihood if there is at least one failure, as long as there is at least one censored observation greater than that failure. It is, however, possible that the maximization algorithm will fail in such cases because the shape of the likelihood can be poorly behaved. For the Bearing Cage example, the probability of obtaining a bootstrap sample with 0 or 1 failures using the resampling method is 0.017 based on a simple binomial distribution computation. Using the FRW method, the probability is 0!
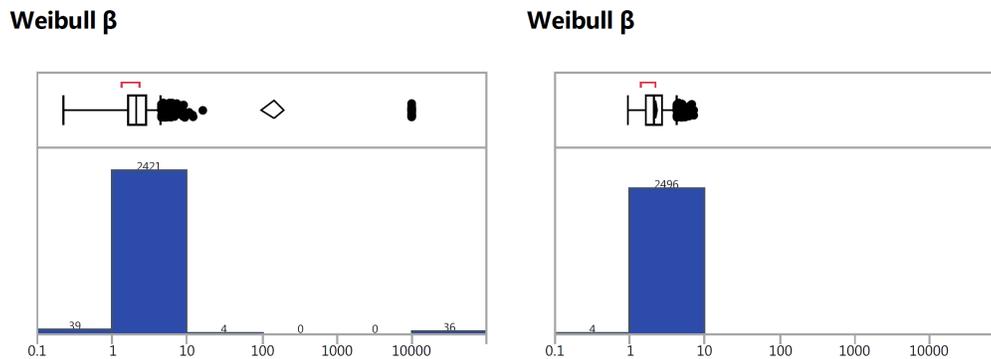
Figure 2: Resampling (left) and fractional-random-weight (right) bootstrap results for the Weibull shape parameter for the bearing cage field-failure data.

Table 3: Resampling and FRW bias-corrected (BC) percentile bootstrap results for the Weibull shape parameter for the bearing cage field-failure data.

| Resampling | | | Fractional-Random-Weight | | |
| Bootstrap Confidence Limits | | | Bootstrap Confidence Limits | | |
Confidence Level	BC Lower	BC Upper	Confidence Level	BC Lower	BC Upper
0.95	1.035	6.540	0.95	1.188	4.402
0.90	1.153	4.866	0.90	1.270	3.898
0.80	1.304	3.750	0.80	1.384	3.344
0.50	1.574	2.671	0.50	1.625	2.641

4.1.3 Bootstrap Results

Figure 2 shows results from the resampling (left) and the FRW bootstrap for the Weibull shape parameter β. The histogram on the left shows that there were 36 samples that resulted in a wild estimate of β which were probably caused by having resamples with 0 failures. The upper endpoint of the confidence interval is even more extreme than that provided by the untrustworthy Wald method. The histogram on the right, based on the FRW bootstrap method is better behaved and the upper endpoint only 4.4. This is a more trustworthy value and is consistent with common experience with fatigue failures in the field. Interestingly (but not surprisingly) the FRW method runs somewhat faster than the resampling method for this example. This is because with the FRW method the optimization algorithm is not faced with bootstrap samples that result in poorly behaved likelihoods that require a lot of time trying to find a maximum that does not exist.

11

Table 4: Rocket motor life data (in years since manufacture).

Years	Number of Motors	Years	Number of Motors	Years	Number of Motors
> 1	105	> 8	211	> 14	14
> 2	164	> 9	124	> 15	5
> 3	153	> 10	90	> 16	3
> 4	236	> 11	72	< 8.5	1
> 5	250	> 12	53	< 14.2	1
> 6	197	> 13	30	< 16.5	1
> 7	230				

4.2 Rocket Motor Field-Failure Weibull Analysis

4.2.1 Background

Olwell and Sorell (2001) present a Bayesian analysis of rocket motor field-failure data. The data were reanalyzed in Chapters 14 and 18 of Meeker, Hahn, and Escobar (2017). There were approximately 20,000 missiles in inventory that had been manufactured over a number of years and put into the stockpile. There had been 1,940 rockets put into flight over a period of time up to 18 years subsequent to their manufacture. At their time of flight, 1,937 of these motors performed satisfactorily; but there were three catastrophic launch failures. The failure probability at 20 years was of interest.

The data are shown in Table 4 where > indicates that the failure time was greater that the indicated year (right-censored observations) and < indicates that the failure time was less than the indicated year (left censored observations). Figure 3(a) is an event plot showing the structure of the data. Although these data could be described with a binary regression model (failure probability as a function of years since manufactured), there are advantages of treating such data as censored failure-time data. In particular, the fraction failing as a function of time is constrained to be monotone increasing and we can use probability plotting to assess whether a chosen distribution is appropriate. (although there is little information for such an assessment in this application.) The usual resampling (integer weight) bootstrap will not work well with this data set, as demonstrated below.

4.2.2 Weibull Analysis

Figure 3(b) is a Weibull probability plot of the rocket motor field failure data. Table 5 gives the numerical ML results and Wald confidence intervals for the distribution parameters. As with the bearing cage example, we will again focus on the Weibull shape parameter to compare
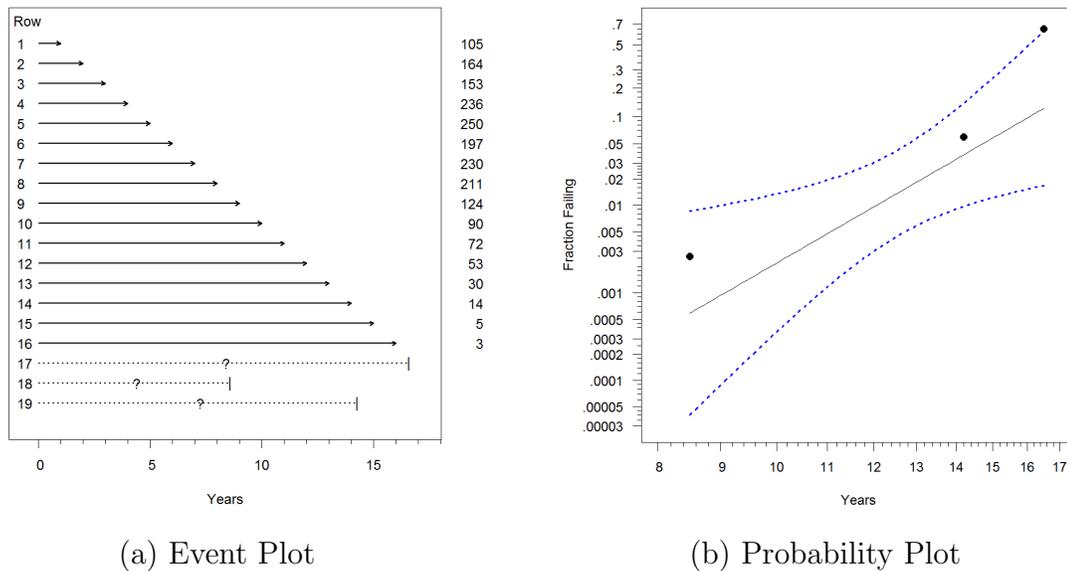
(a) Event Plot　　　　　　　　　　(b) Probability Plot

Figure 3: Event plot and Weibull probability plot for the rocket motor field-failure data.

Table 5: ML estimation results for the Weibull analysis and likelihood-based confidence intervals for the rocket motor field failure data.

Parameter	Estimate	Std Error	95% CI	
			Lower	Upper
η	21.228	4.591	16.846	67.396
β	8.126	3.172	2.963	15.541

the different confidence intervals. The upper endpoint of the Wald confidence interval for β is 34.6. The likelihood interval has an upper endpoint of 15.54, considerably smaller than then that for the Wald interval. Generally, the likelihood interval would be considered to be more trustworthy than the Wald interval. The bootstrap interval would also be expected to provide a trustworthy interval. In this case, however, because of the heavy censoring, the resampling bootstrap method will not work properly.

4.2.3　Bootstrap Results

Figure 4 shows the results for bootstrapping the ML estimate of the Weibull shape parameter for the rocket motor. Using the resampling method there was a large number of samples giving a very large value of the bootstrap estimate of β. These probably arose from bootstrap samples that had no or only one of the left-censored observations in the bootstrap sample. Using the FRW bootstrap, on the other hand resulted in a much more reasonable distribution
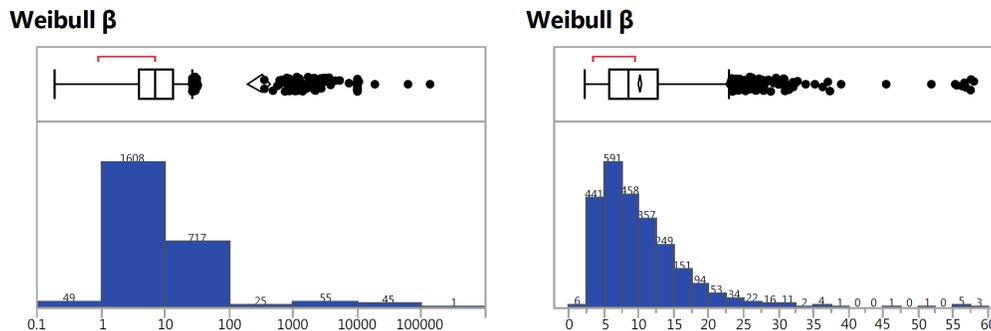
13

Figure 4: Resampling (left) and fractional-random-weight (right) bootstrap results for the Weibull shape parameter for the rocket motor field-failure data.

Table 6: Resampling and FRW bias-corrected (BC) percentile bootstrap results for the Weibull shape parameter β for the rocket motor field-failure data.

Resampling Bootstrap Confidence Limits			Fractional-Random-Weight Bootstrap Confidence Limits		
Confidence Level	BC Lower	BC Upper	Confidence Level	BC Lower	BC Upper
0.95	1.000	10000.000	0.95	2.832	22.912
0.90	1.210	2455.210	0.90	3.097	19.305
0.80	2.296	30.227	0.80	3.666	15.988
0.50	5.348	19.461	0.50	5.207	11.479

and corresponding 95% confidence interval for β. Table 6 shows the resampling and fractional-random-weight bootstrap results for the Weibull shape parameter for the rocket motor field-failure data.

4.3 Bootstrapping the Generalized Gamma Distribution Model for the Ball Bearing Failure Time Data

4.3.1 Background

Meeker and Escobar (1998) and Lawless (2003) fit the generalized gamma distribution to ball bearing life test data that were originally reported in Lieblein and Zelen (1956). Figure 5(a) is an event plot of the data. There was no censoring. The generalized gamma distribution is interesting in that depending on the value of the shape parameter λ, the Weibull ($\lambda = 1$), lognormal ($\lambda = 0$), and Frechet ($\lambda = -1$) distributions are special cases. The detailed
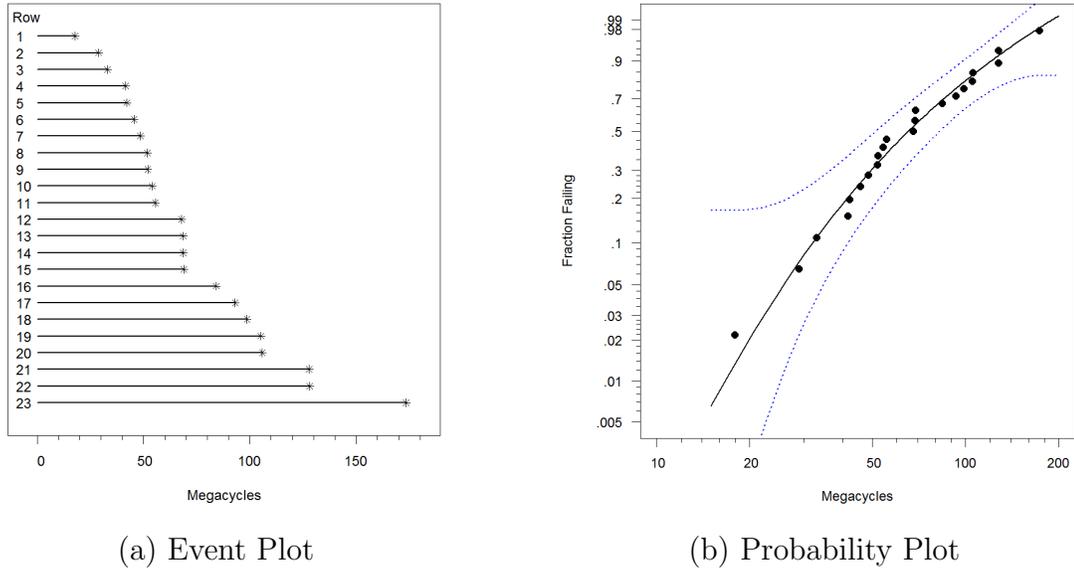
14

(a) Event Plot　　　　　　　　(b) Probability Plot

Figure 5: Event plot and Weibull probability plot for generalized gamma fit to the ball bearing life test data.

expression for the cdf of the generalized gamma distribution is as follows.

$$
F(t; \mu, \sigma, \lambda) = \begin{cases} \Phi_{\text{lg}} \left[\lambda \omega + \log \left(\lambda^{-2} \right), \lambda^{-2} \right], & \text{if } \lambda > 0 \\ \Phi_{\text{nor}}(\omega), & \text{if } \lambda = 0 \\ 1 - \Phi_{\text{lg}} \left[\lambda \omega + \log(\lambda^{-2}), \lambda^{-2} \right], & \text{if } \lambda < 0, \end{cases}
$$

where $t > 0$, $\omega = [\log(t) - \mu] / \sigma$, μ is a location parameter, σ is a scale parameter, and λ is a shape parameter. Here $\Phi_{\text{lg}}(z, \kappa) = \Gamma_{\text{I}}[\exp(z), \kappa]$, $\phi_{\text{lg}}(z, \kappa) = \exp[\kappa z - \exp(z)] / \Gamma(\kappa)$, $\Gamma_{\text{I}}(v, \kappa) = \int_0^v x^{\kappa-1} \exp(-x) dx / \Gamma(\kappa)$ is the incomplete gamma function, and $\Gamma(z)$ is the gamma function.

4.3.2 Maximum Likelihood Estimation

Figure 5(b) shows the ML estimates of the cdf of the generalized gamma distribution, plotted on Weibull probability paper. We can see that it generally fits the data well. Figure 6 shows ML estimates of the Weibull, generalized gamma, and lognormal cdfs on Weibull probability paper. Table 7 gives the ML estimates and Wald confidence intervals for the parameters. The 95% likelihood-based confidence interval for λ is $[-0.76, 1.53]$, somewhat wider than the Wald confidence interval. Again, the likelihood-based interval is generally more trustworthy. In either case, the confidence interval for λ provides an indication that both the lognormal and

15

Table 7: ML estimation results based on generalized gamma distribution and Wald confidence intervals for the ball bearing failure time data.

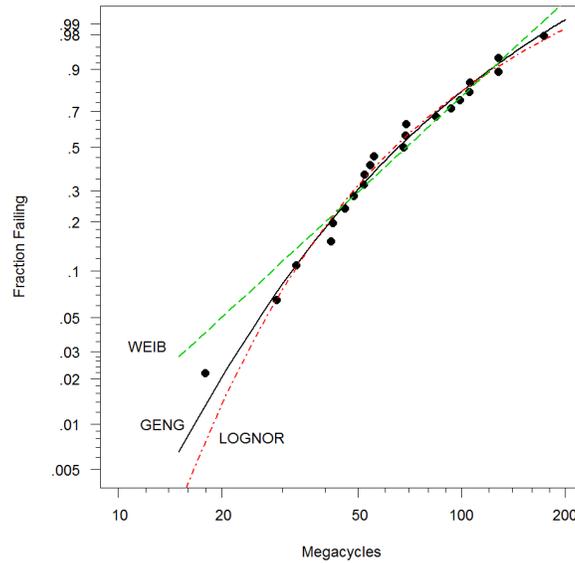| Parameter	Estimate	Std Error	95% CI	
			Lower	Upper
μ	4.230	0.177	3.883	4.577
σ	0.510	0.079	0.354	0.665
λ	0.308	0.549	-0.760	1.383

Figure 6: Ball bearing failure time data fitted by the Weibull, generalized gamma, and log-normal distributions on the Weibull probability paper.

Weibull distributions are consistent with the data. This is because both 0 and 1 lie inside the confidence interval.

The bootstrap provides another alternative to computing confidence intervals for this distribution. In the next section, we will compare the resampling and the FRW bootstrap methods.

4.3.3 Bootstrap Results

Figure 7 gives bootstrap results for the ball bearing generalized gamma distribution shape parameter λ. Table 8 shows the resampling and fractional-random-weight bootstrap results for the generalized gamma distribution shape parameter. The histogram on the left of Figure 7
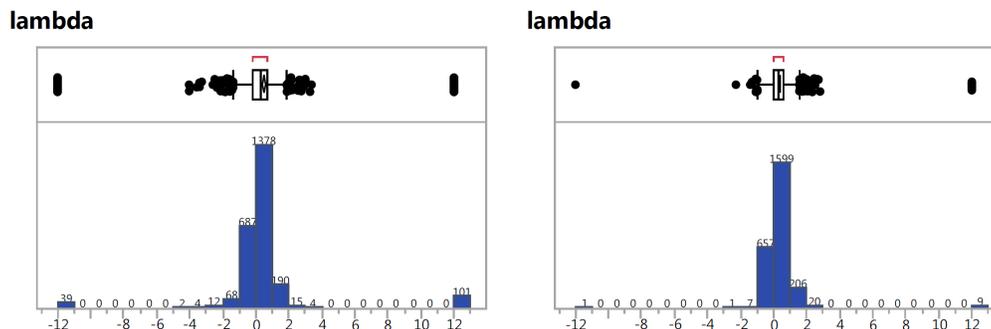
Figure 7: Resampling (left) and fractional-random-weight (right) bootstrap results for the generalized gamma distribution shape parameter for the ball bearing data.

Table 8: Resampling and FRW bias-corrected (BC) percentile bootstrap results for the generalized gamma distribution shape parameter for the ball bearing data.

	Resampling Bootstrap Confidence Limits			Fractional-Random-Weight Bootstrap Confidence Limits	
Coverage	BC Lower	BC Upper	Coverage	BC Lower	BC Upper
0.95	-1.318	12.000	0.95	-0.595	1.704
0.90	-0.877	12.000	0.90	-0.433	1.346
0.80	-0.539	1.243	0.80	-0.263	1.039
0.50	-0.072	0.729	0.50	0.0106	0.642

shows results for the resampling bootstrap method; the histogram on the right shows results using the FRW bootstrap method. The ML algorithm used here constrains λ to be in the interval $[-12, 12]$. With the resampling method, there was ML estimate convergence problems with a substantial number of the bootstrap samples. In particular, there were 39 times the endpoint of the 95% bootstrap interval of the λ parameter was -12 and 101 times where the upper endpoint was $+12$. The 101 values at $+12$ were more than enough to cause the upper endpoint of the 95% bootstrap confidence interval to be $+12$. The corresponding numbers when using the FRW bootstrap method were 1 and 9 times, respectively and these numbers are so small that they have little or no effect on the bootstrap confidence intervals for λ. The FRW method provides a better method for computing confidence intervals for the generalized gamma distribution when the sample size is not large.

5 An Application to Prediction Intervals

In this section, we illustrate the use of FRW in the construction of prediction intervals.
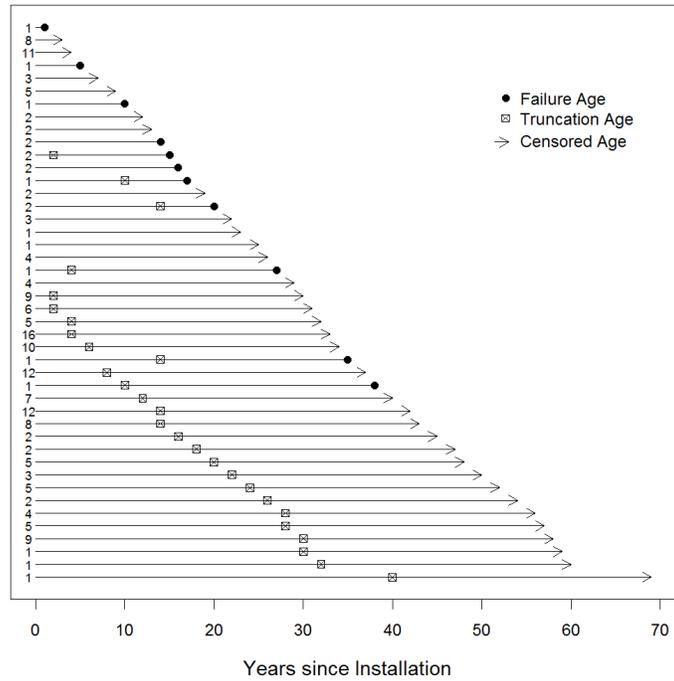
Figure 8: Event plot for the power transformer field-failure data.

5.1 Background

Extending previous work of Escobar and Meeker (1999) and Lawless and Fredette (2005), Hong, Meeker, and McCalley (2009) describe the use of the FRW bootstrap to generate prediction intervals for the number of power transformers that will need to be replaced in future years. The dataset contained information on 710 power transformers with 62 units that had failed. Units still in service at the data freeze date in March 2008 are right censored. Some units that were still in service were more than 60 years old. One difficulty with the data is that records of transformers removed from service before 1980 were not available. Thus units that had been installed before 1980 and which were still in service are observations from a truncated distribution. Figure 8 is an event plot of a representative subset of the data.

There were several categorical covariates, including manufacturer and cooling method that had an effect on the life distribution. It was also learned that, even after adjustment for the other covariates, that there was an important difference between the failure-time distributions of transformers manufactured before and after the mid-1980s. Transformers manufactured before the mid-1980s tend to have longer lifetimes, due to over-engineering that was practiced then.

18

5.2   Modeling and Maximum Likelihood Estimation

The Weibull and lognormal distributions were fit to the data using maximum likelihood. Stratification was based on whether unites were manufactured before or after 1987. The likelihood function is

$$
\begin{aligned}
L(\boldsymbol{\theta}|\text{DATA}) \;=\; & \prod_{i=1}^{n} f(t_i; \boldsymbol{\theta})^{\delta_i \nu_i} \times \left[\frac{f(t_i; \boldsymbol{\theta})}{1 - F(\tau_i^L; \boldsymbol{\theta})} \right]^{\delta_i(1-\nu_i)} \\
& \times \left[1 - F(t_i; \boldsymbol{\theta}) \right]^{(1-\delta_i)\nu_i} \times \left[\frac{1 - F(t_i; \boldsymbol{\theta})}{1 - F(\tau_i^L; \boldsymbol{\theta})} \right]^{(1-\delta_i)(1-\nu_i)} .
\end{aligned}
$$

where t_i is the failure or censoring time, τ_i^L is the lower truncation time, and δ_i and ν_i are censoring and truncation indicators respectively for transformer i. We use $\boldsymbol{\theta}$ to represent the vector of parameters, and $f(t; \boldsymbol{\theta})$ and $F(t; \boldsymbol{\theta})$ are the pdf and cdf of the Weibull distribution, respectively. An important question was how to generate bootstrap samples to do the calibration of the needed prediction intervals. The commonly-used parametric bootstrap would be complicated to implement because it would require a model for the censoring and truncation processes. The resampling method would also have difficulties because of the categorical covariates and the small number of failures in some of the categories. The FRW bootstrap offered an attractive, easy-to-implement alternative that worked without any problems.

5.3   Prediction Results

As an illustration, Figure 9 shows predictions and prediction intervals for the cumulative number of transformer failures for the next ten years, starting in 2008. The prediction is made for transformers that were installed after 1987. The predicted number failing for the latter group is larger both because the lifetimes of the newer models tend to be shorter and also because the size of the risk set was larger.

Figure 10 shows, for a subset of the transformers that were still in operation at the time the predictions were made, the age of the transformer and a prediction interval quantifying the information available about the distribution of remaining life for individual transformers. Although some of the upper endpoints of the prediction intervals are likely overly optimistic (probably because they rely on extrapolation), the lower endpoints allowed a useful ranking of which transformers where at highest risk for failure in the short term.
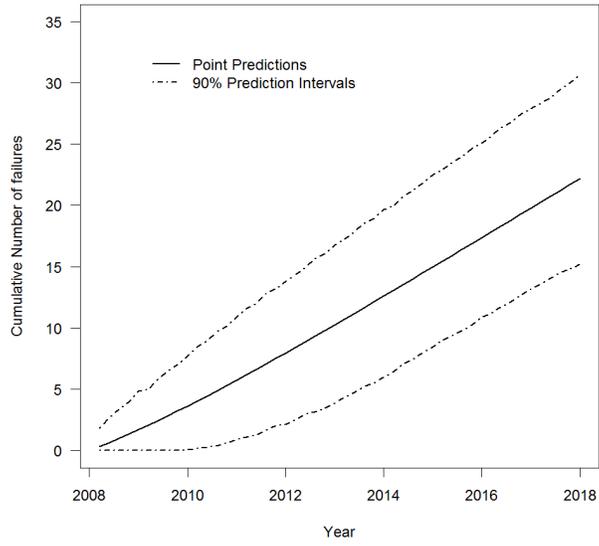
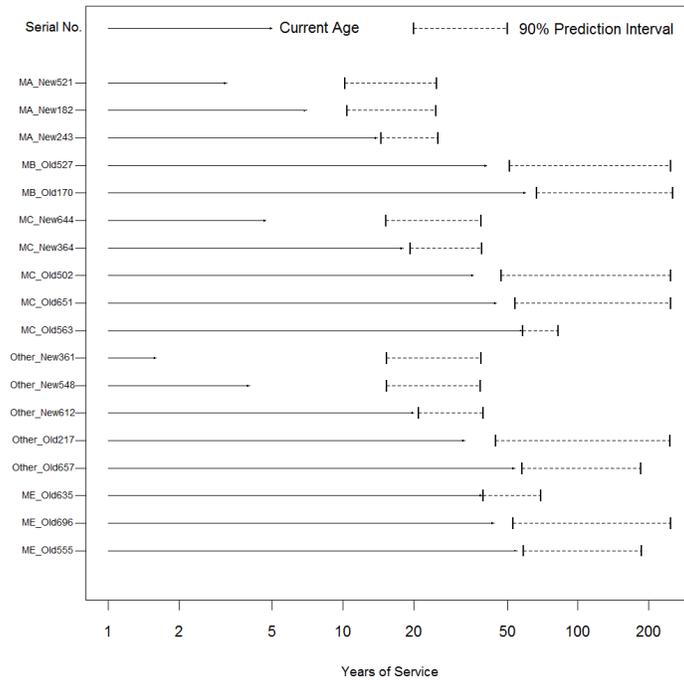Figure 9: Power transformer fleet predictions based on the fractional-random-weight bootstrap.

Figure 10: Power transformer individual predictions based on the fractional-random-weight bootstrap.

6 An Application in Design of Experiments

6.1 Background

Design of experiments is a common approach to problem-solving in science and industry. Designed experiments are specially structured to obtain as much information in as few samples as possible. They are usually so efficient that the removal of even small numbers of observations can induce model singularities that fundamentally change the meaning of the estimated parameters in unpredictable ways. For this reason, the resampling bootstrap is generally avoided in the analysis of designed experiments. An often-stated goal is to obtain as much information as possible about the relationship between the experimental factors (x) and the response variable(s) (y). Usually, a designed experiment uses a specially constructed combination of x values that optimize information gained in a small number of runs. After the data become available, then there is a need to decide on the appropriate statistical model to describe the relationship between x and y.

6.2 Using the Bootstrap in Model Selection

The bootstrap is a useful tool for identifying the subset of the x variables (as well as possible interaction and quadratic effects) that best explain variation in y. The resampling bootstrap, however, can encounter problems because the removal of observations can drastically change which parameters can be estimated. There are two well-known alternatives to resampling:

- Using a parametric bootstrap (simulating data from a given model), and
- Resampling residuals from a fitted model.

The problem with these two methods is that they require specifying a model, which is what we are trying to determine! As mentioned in Section 2.2, the FRW bootstrap can keep all observations during the modeling process, and thus suitable for model-building applications that can be used with data from a designed experiment.

6.3 The Nitrogen Oxides Example

Nitrogen Oxides (NOx) are toxic greenhouse gases that are common by-products of burning organic compounds. An experiment was done on an industrial burner to study the amount of NOx it created. A 32 run I-Optimal response surface model (RSM) design was created with 7 continuous factors:

- Hydrogen Fraction in primary fuel
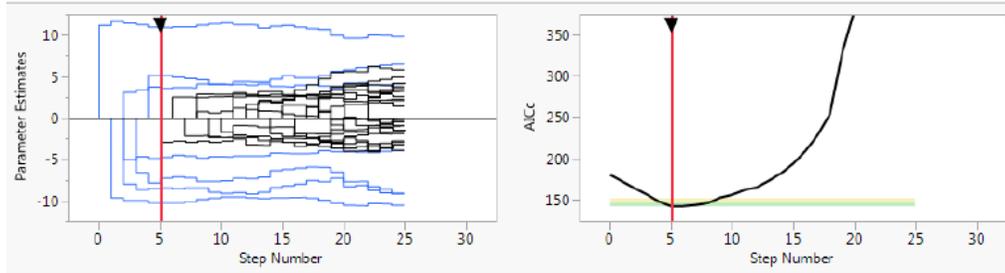- Air/Fuel Ratio
- Lance Position X

Figure 11: Results for model selection using forward stepwise selection.

- Lance Position Y
- Secondary Fuel Fraction
- Dispersant
- Ethanol Percentage in primary fuel

This design would allow estimation of all main effects, two-factor interactions and quadratic effects. We want to assess the importance of the input variables (including the two-factor interactions and quadratic terms).

6.4   Using Forward Selection

First, we apply a forward stepwise procedure that selects a model using the AIC model selection criterion. The results are shown in Figure 11 and Table 9. To better understand the stability of this model choice and to explore the possibility that other variables might make an important contribution, it is possible to apply the FRW bootstrap method to the model building procedure. Then the results of such a bootstrap can be used to obtain selection probabilities for the different model terms.

6.5   Bootstrapping the Forward Selection Procedure

We use the FRW approach to bootstrap the forward selection procedure. One thousand FRW bootstrap data sets were generated. For each FRW bootstrap data set, the forward selection procedure is applied and the corresponding row in the table gives the values of the regression coefficients. The zeros in the table indicate that the variable was not included in the model for that bootstrap sample.

Table 10 shows partial results for the first 16 FRW bootstrap data sets.(i.e., for some of the regression coefficients). Figure 12 shows the histograms that summarize the FRW bootstrap modeling results. The spikes at 0 in some of the histograms indicate the number of times that the corresponding variable did not enter the model (frequently for Lance Position $Y(5, 10)$

22

Table 9: Results from using forward stepwise selection to choose a model, showing the parameter estimates for the original predictors.

| Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	95% CI Lower	95% CI Upper
Intercept	30.305	0.431	4939.196	<0.0001	29.459	31.150
Hydrogen Fraction$(0, 6)$	2.454	0.341	51.715	<0.0001	1.785	3.123
Air/Fuel Ratio$(1.05, 1.3)$	−2.3030	0.344	44.855	<0.0001	−2.977	−1.629
Lance Position $X(21, 25)$	0.853	0.305	7.800	<0.0052	0.254	1.451
Lance Position $V(5, 10)$	0	0	0	1	0	0
Sec. Fuel Fraction$(0, 6)$	−1.101	0.261	17.817	<0.0001	−1.612	−0.590
Dispersant$(0.5, 1.5)$	0	0	0	1	0	0
Ethanol$(0, 10)$	0	0	0	1.000	0	0
Hydrogen Fraction	0	0	0	1	0	0
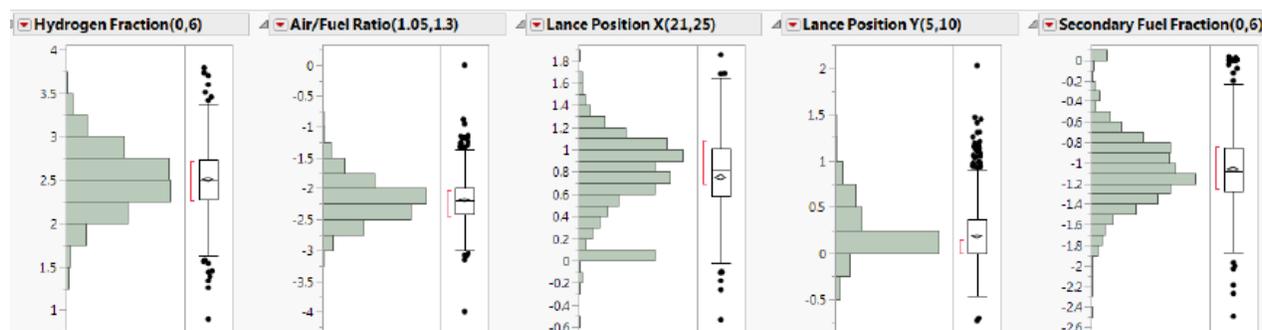
Figure 12: Histograms that summarize the FRW Bootstrap modeling results.

and never for Hydrogen Fraction$(0, 6)$).

The results in Table 10 can be used to construct a table like that in Table 11. This table gives the proportion of times across the 1000 bootstrap samples that each variable was chosen to be in the model. Because of the Bayesian bootstrap result given in Rubin (1981), these proportions can be interpreted as the posterior probability that the corresponding model term should be in the model. One could then use a cutoff point (such as 0.50) to decide whether model terms should be included or not.

7 Concluding Remarks and Areas for Future Research

With vastly improved computing capabilities and the bootstrap theory that has been developed over the past 40 years, bootstrapping provides an important useful tool for obtaining confidence intervals, prediction intervals, and better regression models.

Table 10: Table showing a snapshot of the bootstrap estimates for the first 16 FRW bootstrap data sets (only the estimates of the first five variables are shown).

Hydrogen Fraction $(0, 6)$	Air/Fuel Ratio $(1.05, 1.3)$	Lance Position X $(21, 25)$	Lance Position Y $(5, 10)$	Sec. Fuel Fraction $(0, 6)$
2.450	-2.303	0.853	0	-1.101
2.231	-2.196	0.964	0	-1.137
3.006	-2.242	0.714	0	-1.306
2.657	-1.929	0.915	0.235	-1.154
2.987	-2.546	0.085	0.459	-1.528
2.163	-1.567	0.945	-0.067	-0.761
1.804	-2.007	1.113	0	-1.629
2.830	-2.549	1.061	0.259	-1.198
2.427	-2.035	1.004	0	-1.497
2.487	-1.881	0.345	0	-0.736
2.617	-2.145	0	0.601	-0.731
1.762	-1.098	1.380	0	-1.446
2.416	-2.244	1.075	0.518	-1.105
2.697	-2.058	1.064	0.570	-0.890
3.028	-2.179	1.190	0.252	-1.121
2.649	-2.193	0	0	-1.488

The FRW bootstrap tremendously expands the potential areas of application of the bootstrap to applications involving heavy censoring and/or truncation, categorical explanatory variable, and designed experiments where dropping certain combinations of the original observations can cause estimability problems. Those problems do not arise when the FRW bootstrap is used.

Overall, we observe that the FRW bootstrap is as easy to implement as the resampling bootstrap, and it has similar desirable properties as the resampling bootstrap in situations where the resampling bootstrap works well. The FRW bootstrap also retains desirable properties even when the resampling bootstrap breaks down. Through the examples in this paper, we see that the FRW bootstrap as a safer, more broadly applicable, alternative to the resampling bootstrap.

There are a number of areas that could be investigated to provide further insight into when and how the FRW bootstrap methods should be used with finite samples.

• There are different, asymptotically equivalent ways to choose the random weighs for bootstrapping (including resampling). This leaves open the question about differences in the properties of bootstrap procedures in finite samples. For example, if weights are chosen to have a mean and variance of one, what would be the effect on the performance of varying the third or higher moments?

Table 11: The proportion of times across the 1000 FRW bootstrap samples that each variable was chosen to be in the model.

Term	Proportion Selected
Hydrogen Fraction$(0, 6)$	1.000
Air/Fuel Ratio$(1.05, 1.3)$	0.999
Secondary Fuel Fraction$(0, 6)$	0.981
Air/Fuel Ratio $*$ Air/Fuel Ratio	0.946
Lance Position $X(21, 25)$	0.903
Lance Position X $*$ Secondary Fuel Fraction	0.848
Hydrogen Fraction $*$ Secondary Fuel Fraction	0.733
Lance Position $Y(5, 10)$	0.593
Dispersant$(0.5, 1.5)$	0.551
Secondary Fuel Fraction $*$ Secondary Fuel Fraction	0.479
Lance Position Y $*$ Lance Position Y	0.294
Hydrogen Fraction $*$ Lance Position Y	0.224
Hydrogen Fraction $*$ Hydrogen Fraction	0.203
Lance Position Y $*$ Dispersant	0.192
Hydrogen Fraction $*$ Lance Position X	0.180
Air/Fuel Ratio $*$ Lance Position Y	0.177
Air/Fuel Ratio $*$ Dispersant	0.163
Ethanol$(0, 10)$	0.113
Hydrogen Fraction $*$ Air/Fuel Ratio	0.104
Lance Position X $*$ Lance Position X	0.083

• We have demonstrated a clear advantage for the FRW bootstrap in situations where estimability problems occur when certain combinations of observations are dropped. In situations where there will be no estimability problems is it possible that the FRW approach has other advantages. It would be useful to compare different nonparametric and parametric methods for generating bootstrap estimates when using a parametric model to describe one's data. In particular, it would be interesting to compare, resampling methods, a fully parametric bootstrap simulation (e.g., where the censoring distribution is modeled), and FRW bootstrap to see if there are important differences in bootstrap performance.

• Generalized fiducial inference (GFI), which is also known as generalized pivotal quantity (see Hannig, Iyer, and Patterson 2006, Majumder and Hannig 2016, and Hannig et al. 2016), has proven to be a powerful tool for defining confidence interval procedures for non-standard models. Implementing GFI methods generally requires computing a large set of simulated parameter estimates, in a manner that is similar to the parametric bootstrap. In situations involving heavy censoring, even the parametric bootstrap sampling will have estimability problems. Use of FRW instead should allow GFI methods to be used in such applications.

Acknowledgments

A   Proof of Theoretical Results

In this appendix, we prove the theoretical results given in Section 3.

A.1   Result 1: Consistency

Let $h(\boldsymbol{\theta}) = \mathbf{E}[l_i(\boldsymbol{\theta}; X_i)]$, $i = 1, \ldots, n$. Note that $\mathbf{E}[\bar{l}(\boldsymbol{\theta})] = h(\boldsymbol{\theta})$. Also, $\mathrm{Var}[\bar{l}(\boldsymbol{\theta})] \to 0$, as $n \to \infty$, because $\widehat{\boldsymbol{\theta}}$ is consistent. The expectation of $\bar{l}^*(\boldsymbol{\theta})$ is

$$\mathbf{E}\left[\bar{l}^*(\boldsymbol{\theta})\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\left[Z_i l_i(\boldsymbol{\theta}; X_i)\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\left\{\mathbf{E}\left[Z_i l_i(\boldsymbol{\theta}; X_i)\right]|X_i\right\} \tag{3}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\left[l_i(\boldsymbol{\theta}; X_i)\right] = h(\boldsymbol{\theta}).$$

In (3), two layers of expectation are involved, in which the inner expectation is taken with respect to Z_i given X_i and the outer expectation is taken with respect to X_i. In addition,

$$\text{Var}\left[\bar{l}^*(\boldsymbol{\theta})\right] = \text{Var}\left\{\mathbf{E}\left[\bar{l}^*(\boldsymbol{\theta})|\boldsymbol{X}\right]\right\} + \mathbf{E}\left\{\text{Var}[\bar{l}^*(\boldsymbol{\theta})|\boldsymbol{X}]\right\} \quad (4)$$
$$= \frac{1}{n}[2\text{Var}[l_i(\boldsymbol{\theta}, X_i)] + h(\boldsymbol{\theta})^2] \rightarrow 0,$$

as $n \rightarrow \infty$. Similarly, for the variance/expectation pairs in (4), the inside operator is taken with respect to Z_i given X_i, and the outside operator is taken with respect to X_i. Thus, $\bar{l}^*(\boldsymbol{\theta})$ is consistent for $h(\boldsymbol{\theta})$, and $\bar{l}(\boldsymbol{\theta})$ and $\bar{l}^*(\boldsymbol{\theta})$ converge to the same function $h(\boldsymbol{\theta})$. Equivalently, the solutions of $\bar{l}'(\boldsymbol{\theta}) = 0$ and $\bar{l}^{*'}(\boldsymbol{\theta}) = 0$ converge to $\boldsymbol{\theta}$, proving **Result 1**.

A.2   Result 2: Asymptotic Normality

Consider the following Taylor series expansion,

$$0 = \bar{l}^{*'}(\widehat{\boldsymbol{\theta}}^*) = \bar{l}^{*'}(\widehat{\boldsymbol{\theta}}) + \bar{l}^{*''}(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}}) + \cdots .$$

Thus,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})|\boldsymbol{X}_n \approx -[\bar{l}^{*''}(\widehat{\boldsymbol{\theta}})]^{-1}[\sqrt{n}\,\bar{l}^{*'}(\widehat{\boldsymbol{\theta}})]. \quad (5)$$

Here $\bar{l}^{*''}(\boldsymbol{\theta})$ is the Hessian matrix. In the following, we will show that, conditional on $\boldsymbol{X}_n$, $\bar{l}^{*''}(\boldsymbol{\theta})$ converges to the Fisher information matrix $I(\boldsymbol{\theta})$, and $\sqrt{n}\,\bar{l}^{*'}(\widehat{\boldsymbol{\theta}}) \rightarrow \text{N}[0, \sigma^2 I(\boldsymbol{\theta})]$, as $n \rightarrow \infty$. Note that $\sum_{i=1}^{n} l_i'(\widehat{\boldsymbol{\theta}}) = 0$. Thus, $\sqrt{n}\,\bar{l}^{*'}(\widehat{\boldsymbol{\theta}})$ can be written as

$$\sqrt{n}\,\bar{l}^{*'}(\widehat{\boldsymbol{\theta}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i l_i(\boldsymbol{\theta}; X_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - 1)l_i'(\widehat{\boldsymbol{\theta}}).$$

In the following, we use $\mathbf{E}_*$ and Var_* to denote bootstrap expectation and variance. That is, these are moments determined by the iid exponential variables $Z_1, \ldots, Z_n$ treating the data $\boldsymbol{X}_n$ as fixed. The bootstrap mean and variance of $\sqrt{n}\,\bar{l}^{*'}(\widehat{\boldsymbol{\theta}})$ are obtained as

$$\mathbf{E}_*\left[\sqrt{n}\,\bar{l}^{*'}(\widehat{\boldsymbol{\theta}})\right] = \mathbf{E}_*\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - 1)l_i'(\widehat{\boldsymbol{\theta}})\right] = 0,$$

and

$$\text{Var}_*\left[\sqrt{n}\,\bar{l}^{*'}(\widehat{\boldsymbol{\theta}})\right] = \text{Var}_*\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - 1)l_i'(\widehat{\boldsymbol{\theta}})\right] = \frac{1}{n}\sum_{i=1}^{n}\left[l_i'(\widehat{\boldsymbol{\theta}})\right]^2,$$

respectively. Note that $\sum_{i=1}^{n}\left[l_i'(\widehat{\boldsymbol{\theta}})\right]^2/n \rightarrow I(\boldsymbol{\theta})$. That is, the sample mean of squared score functions at the ML estimate converges in probability to the information matrix $I(\boldsymbol{\theta})$.

Hence, by using the weighted central limit theorem (i.e., the sum of independent, but weighted, exponential variables $Z_1, \ldots, Z_n$, given $\boldsymbol{X}_n$),

$$\sqrt{n}\,\bar{l}^{*\prime}(\widehat{\boldsymbol{\theta}}) \to \mathrm{N}\left[0, \sigma^2 I(\boldsymbol{\theta})\right], \tag{6}$$

in which the convergence is in bootstrap probability given $\boldsymbol{X}_n$. Also,

$$\bar{l}^{*\prime\prime}(\widehat{\boldsymbol{\theta}}) \to I(\boldsymbol{\theta}). \tag{7}$$

in which the convergence is in bootstrap probability. That is because, conditional on $\boldsymbol{X}_n$, $\bar{l}^{*\prime\prime}(\widehat{\boldsymbol{\theta}})$ has mean $\bar{l}''(\widehat{\boldsymbol{\theta}})$ and variance $\sum_{i=1}^n [l_i''(\boldsymbol{\theta})]^2/n^2$ which goes to zero. Note that $\bar{l}''(\widehat{\boldsymbol{\theta}})$ converges in probability to $I(\boldsymbol{\theta})$.

By (6), (7), and the Slutsky's theorem, we have $\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})|\boldsymbol{X}_n \to \mathrm{N}[0, I(\boldsymbol{\theta})^{-1}]$, proving **Result 2**.

A.3 Result 3: Existence of ML Estimates

Here we provide a proof of the result for the Weibull distribution. The proof of other commonly used log-location-scale distributions (i.e., lognormal, Fréchet, and loglogistic distributions) is similar but lengthy, and thus it is omitted here. Here we prove the result under two particular cases. The pdf of the Weibull distribution is

$$f(t; \eta, \beta) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \; t > 0.$$

For case (i) with two distinct failure times t_1 and t_2, we suppose that $t_1 < t_2$, and w_1 and w_2 are the weights such that $w_1 + w_2 = 1$. The weighted loglikelihood under the Weibull distribution is

$$l(\eta, \beta) = \sum_{i=1}^{2} w_i \log\left[f\left(t_i; \eta, \beta\right)\right] = \sum_{i=1}^{2} w_i \left[\log(\beta) + (\beta - 1)\log(t_i) - \beta\log(\eta) - \left(\frac{t_i}{\eta}\right)^{\beta}\right].$$

Taking the derivative with respect to η and setting it to 0, we obtain

$$\frac{\partial l}{\partial \eta} = \sum_{i=1}^{2} w_i \left(-\frac{\beta}{\eta} + \frac{\beta t_i^{\beta}}{\eta^{\beta+1}}\right) = 0. \tag{8}$$

Simplifying (8), we obtain

$$-\frac{\beta}{\eta} + \frac{\sum_{i=1}^{2} w_i \beta t_i^{\beta}}{\eta^{\beta+1}} = 0.$$

Solving the above equation, we obtain $\widehat{\eta} = \left(\sum_{i=1}^{2} w_i t_i^{\beta} \right)^{\frac{1}{\beta}}$. Substituting $\widehat{\eta}$ into the loglikelihood, we obtain the log of the profile likelihood for β as follows,

$$\widehat{l}(\beta) = l(\widehat{\eta}, \beta) = \sum_{i=1}^{2} w_i \left[\log(\beta) + (\beta - 1)\log(t_i) - \log\left(\sum_{i=1}^{2} w_i t_i^{\beta} \right) - \frac{t_i^{\beta}}{\sum_{i=1}^{2} w_i t_i^{\beta}} \right].$$

Taking the derivative with respect to β, we obtain,

$$\frac{\partial \widehat{l}}{\partial \beta} = \frac{1}{\beta} + \sum_{i=1}^{2} w_i \log(t_i) - K,$$

where $K = \left[\sum_{i=1}^{2} w_i \log(t_i) t_i^{\beta} \right] / \left(\sum_{i=1}^{2} w_i t_i^{\beta} \right)$. When $\beta \to 0$, $1/\beta \to \infty$, and the term K is finite. Hence, $\partial \widehat{l} / \partial \beta \to \infty$. When $\beta \to \infty$, $1/\beta \to 0$, and the term

$$K \to \frac{w_2 \log(t_2) t_2^{\beta}}{w_2 t_2^{\beta}} = \log(t_2).$$

We obtain,

$$\frac{\partial \widehat{l}}{\partial \beta} \to w_1 \log(t_1) + w_2 \log(t_2) - \log(t_2) = w_1[\log(t_1) - \log(t_2)],$$

where $w_1[\log(t_1) - \log(t_2)]$ is a negative constant. Thus, $\partial \widehat{l} / \partial \beta = 0$ has at least one root.

To further show that $\widehat{l}$ has a global maximum, we restrict β to a closed interval. Because $\lim_{\beta \to 0} \partial \widehat{l} / \partial \beta = \infty$, let $\tilde{\beta}_{\text{low}} = \min\{\beta | \beta \geq 0, \partial \widehat{l} / \partial \beta \leq C\}$ where C is a large positive constant. Then the global maximum $\widehat{\beta}$ must be greater than or equal to $\tilde{\beta}_{\text{low}}$ because in the interval $(0, \tilde{\beta}_{\text{low}})$, $\widehat{l}$ is increasing. Because $\widehat{l}$ is continuous and $\lim_{\beta \to \infty} \partial \widehat{l} / \partial \beta = w_1[\log(t_1) - \log(t_2)]$, there exists $\tilde{\beta}_{\text{high}}$ such that $\forall \beta \geq \tilde{\beta}_{\text{high}}, \partial \widehat{l} / \partial \beta \leq 0.5 w_1[\log(t_1) - \log(t_2)]$. Then $\widehat{\beta}$ must be less than or equal to $\tilde{\beta}_{\text{high}}$ because $\widehat{l}$ is decreasing when $\beta \geq \tilde{\beta}_{\text{high}}$. Then we have $\widehat{\beta} \in [\tilde{\beta}_{\text{low}}, \tilde{\beta}_{\text{high}}]$. Because $\widehat{l}$ is continuous in a bounded interval $[\tilde{\beta}_{\text{low}}, \tilde{\beta}_{\text{high}}]$, it must have a maximum in the closed interval, proving case (i).

For case (ii) with one failure time t_1 and a right-censored observation t_2 such that $t_2 > t_1$, the weighted loglikelihood under the Weibull distribution is

$$l(\eta, \beta) = w_1 \log[f(t_1, \eta, \beta)] + w_2 \log[1 - F(t_2, \eta, \beta)]$$
$$= w_1 \left[\log(\beta) + (\beta - 1)\log(t_1) - \beta \log(\eta) - \left(\frac{t_1}{\eta} \right)^{\beta} \right] + w_2 \left[-\left(\frac{t_2}{\eta} \right)^{\beta} \right].$$

Taking the derivative with respect to η and setting it to 0, we obtain,

$$\frac{\partial l}{\partial \eta} = w_1 \left(-\frac{\beta}{\eta} + \frac{\beta t_1^{\beta}}{\eta^{\beta+1}} \right) + w_2 \frac{\beta t_2^{\beta}}{\eta^{\beta+1}} = 0.$$

Solving the above equation, we obtain

$$\widehat{\eta} = \frac{w_1 t_1^{\beta} + w_2 t_2^{\beta}}{w_1}.$$

Substituting $\widehat{\eta}$ into $l(\eta, \beta)$, we obtain the log of the profile likelihood for β as follows,

$$\widehat{l}(\beta) = l(\widehat{\eta}, \beta)$$
$$= w_1 \left[\log(\beta) + (\beta - 1)\log(t_1) - \log\left(\frac{\sum_{i=1}^{2} w_i t_i^{\beta}}{w_1} \right) - \frac{w_1 t_1^{\beta}}{\sum_{i=1}^{2} w_i t_i^{\beta}} \right] + w_2 \left(\frac{w_1 t_2^{\beta}}{\sum_{i=1}^{2} w_i t_i^{\beta}} \right).$$

Differentiating with respect to β gives

$$\frac{\partial \widehat{l}}{\partial \beta} = w_1 \frac{1}{\beta} + w_1 \log(t_1) - w_1 K.$$

Similar to case (i), one we can restrict $\widehat{\beta}$ to a closed interval $[\tilde{\beta}_{\text{low}}, \tilde{\beta}_{\text{high}}]$ and show that $\widehat{\beta}$ exists.

When the sample size n is greater than 2, the proof is similar. The condition for the existence of the ML estimate will change slightly when $n > 2$. For case (i), there should be at least two distinct failures. For case (ii), the condition is that there exists one right-censored observation which exceeds the failure time.

References

Abernethy, R. B., J. E. Breneman, C. H. Medlin, and G. L. Reinman (1983). Weibull analysis handbook. Technical report, Air Force Wright Aeronautical Laboratories, URL: http://www.dtic.mil/dtic/tr/fulltext/u2/a143100.pdf.

Barbe, P. and P. Bertail (1995). *The Weighted Bootstrap*. Springer.

Chatterjee, S. and A. Bose (2005). Generalized bootstrap for estimating equations. *Annals of Statistics 33*, 414–436.

Chiang, C.-T., L. F. James, and M.-C. Wang (2005). Random weighted bootstrap method for recurrent events with informative censoring. *Lifetime Data Analysis 11*, 489–509.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*. London: Chapman and Hall.

Davison, A. C. and D. V. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

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

Escobar, L. A. and W. Q. Meeker (1999). Statistical prediction based on censored life data. *Technometrics 41*, 113–124.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.

Hannig, J., H. Iyer, R. C. S. Lai, and T. C. M. Lee (2016). Generalized fiducial inference: A review and new results. *Journal of the American Statistical Association 111*, 1346–1361.

Hannig, J., H. Iyer, and P. Patterson (2006). Fiducial generalized confidence intervals. *Journal of the American Statistical Association 101*, 254–269.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician 69*, 371–386.

Hong, Y., W. Q. Meeker, and J. D. McCalley (2009). Prediction of remaining life of power transformers based on left truncated and right censored lifetime data. *Annals of Applied Statistics 3*, 857–879.

Jin, Z., Z. Ying, and L. Wei (2001). A simple resampling method by perturbing the minimand. *Biometrika 88*, 381–390.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data (Second Edition)*. John Wiley & Sons.

Lawless, J. F. and M. Fredette (2005). Frequentist prediction intervals and predictive distributions. *Biometrika 92*, 529–542.

Lieblein, J. and M. Zelen (1956). Statistical investigation of the fatigue life of deep-groove ball bearings. *Journal of Research, National Bureau of Standards 57*, 273–316.

Majumder, A. P. and J. Hannig (2016). Higher order asymptotics of generalized fiducial distribution. *arXiv:1608.07186 [math.ST]*.

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

Meeker, W. Q., G. J. Hahn, and L. A. Escobar (2017). *Statistical Intervals: A Guide for Practitioners and Researchers*. John Wiley & Sons.

Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B 56*, 3–48.

Olwell, D. H. and A. A. Sorell (2001). Warranty calculations for missiles with only current-status data, using Bayesian methods. In *Annual Reliability and Maintainability Symposium. 2001 Proceedings. International Symposium on Product Quality and Integrity (Cat. No.01CH37179)*, pp. 133–138.

Rubin, D. B. (1981). The Bayesian bootstrap. *Annals of Statistics 9*, 130–134.

Shao, J. and D. Tu (1995). *The Jackknife and Bootstrap*. Springer.

Xu, Z., Y. Hong, and W. Q. Meeker (2015). Assessing risk of a serious failure mode based on limited field data. *IEEE Transactions on Reliability 64*, 51–62.