Applications of the Fractional-Random-Weight Bootstrap

Li Xu[1], Chris Gotwalt[2], Yili Hong[1], Caleb B. King[2], and William Q. Meeker[3]

[1]Department of Statistics, Virginia Tech, Blacksburg, VA 24061

[2]JMP Division, SAS, Research Triangle, NC 12345

[3]Department of Statistics, Iowa State University, Ames, IA 50011

Abstract

For several decades, the resampling based bootstrap has been widely used for computing confidence intervals (CIs) for applications where no exact method is available. However, there are many applications where the resampling bootstrap method can not be used. These include situations where the data are heavily censored due to the success response being a rare event, situations 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. These three situations all have in common that there may be a substantial proportion of the resamples where it 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 CIs in a way that is accessible to statistical practitioners. The fractional-random-weight bootstrap method is easy to use and has advantages over the resampling method in many challenging applications.

Key Words: Bayesian bootstrap, Censored data, Confidence interval, Prediction interval, Random weighted bootstrap, Variable selection.

1 Introduction

1.1 Bootstrap Background

The bootstrap is a popular statistical tool used to obtain inferences, such as approximate confidence intervals (CIs) and approximate prediction intervals that have coverage probabilities close to the nominal confidence level. Bootstrapping is a set of procedures for sampling from the distribution of an estimator, employing various data generation and augmentation procedures to create new datasets from which new individual values of the estimator are computed. These estimated distributions of the estimators can then be used for many purposes, including creating 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 can be applied even in situations where classical theory offers little or no guidance on how to compute CIs. Generally, there are only minimal regularity conditions (such as a finite variance and a certain degree of smoothness) needed to make bootstrap methods work well. Technical details of bootstrap methods can be found in classical references such as Hall (1992), Lo (1993), 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 two categories: nonparametric and parametric. Nonparametric bootstrap procedures require no assumptions about the shape of the underlying data-generating probability distribution. The most common approach is to generate a sequence of new datasets by sampling the rows of the original data with replacement. 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 in the new datasets must be done in a manner that mimics 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 either unknown or are too complicated. In such situations, the nonparametric resampling method is much easier to implement.

After each bootstrap dataset is generated, the statistical procedure (e.g., model fitting, 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., 2,000 times) and then the saved results are processed to make inferences (e.g., construct CIs). There are many different ways to use bootstrap samples to compute a CI (e.g., simple percentile, bias-corrected (BC) percentile, BC and accelerated, percentile-t intervals). In Jeng and Meeker (1999), there are detailed descriptions for these

bootstrap CI constructions.

1.2 The Idea of Data Weights

In many data analysis applications, it is convenient to put weights on observations. Weights are also referred to as frequencies or counts in some cases. In this paper, the weights we consider need not sum to one. There are many examples of counts and weights in different areas. 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, which typically consist of failure times (all having weight 1 except in the case of ties). For those units censored at the same time, the censored data can be summarized into one row by provided the censoring time and the counts of the censored units. Weights are also used when data are binned, where the weights indicate the number of observations in each bin (e.g., as displayed in a histogram). In survey sampling, weights are used to make data more representative of a population, and in causal modeling using propensity methods, weights are used to make the distribution of control observations more similar to the distribution of treatment observations.

The resampling bootstrap method can also be viewed as resampling data with random integer weights (e.g., Efron 1982). That is, each observation has a weight indicating the number of times it was drawn in the resampling. Rubin (1981) introduces the Bayesian bootstrap, which uses all original observations, with non-integer weights on the observations, which is an example of a fractional-random-weight (FRW) bootstrap. We give an explicit example of this in the next section.

Many statistical estimation methods allow the use of weights or frequencies. For example, consider a data vector $(y_1, y_2, \ldots, y_n)'$ with corresponding weights $(w_1, w_2, \ldots, w_n)'$. Then estimates of the mean (μ) and variance (σ^2) can be computed from

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

More generally, suppose we have a data matrix $D = (\boldsymbol{x}_1', \boldsymbol{x}_2', \ldots, \boldsymbol{x}_n')'$ with corresponding weights $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)'$, where each $\boldsymbol{x}_i$ is a row in the data matrix, which may contain information such as a response, explanatory variables, and censoring or truncation indicators for observation i. Then the weighted likelihood is

$$L(\boldsymbol{\theta}; D, \boldsymbol{w}) = \mathcal{C} \prod_{i=1}^n \left[L_i(\boldsymbol{\theta}; \boldsymbol{x}_i) \right]^{w_i}. \tag{1}$$

Here, $\boldsymbol{\theta}$ in (1) is a general notation for the unknown parameters, $\mathcal{C}$ is a constant unrelated to $\boldsymbol{\theta}$, and $L_i(\boldsymbol{\theta}; \boldsymbol{x}_i)$ 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. It also provides an alternative way to understand bootstrap methods. The objective of this paper is to review the FRW bootstrap method and demonstrate, in a way that is highly accessible to statistical practitioners, how to apply it to applications in which the resampling (integer weights) bootstrap methods tend not to 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. An important advantage of the FRW is not having to worry about estimability in these applications, allowing for many new applications for the method.

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), which also compares the small-sample coverage properties of different bootstrap methods.

As there are only a handful of articles devoted to the FRW bootstrap sampling method and it appears to be under-appreciated in spite of its usefulness. We believe the FRW bootstrap method could serve a much larger role in the toolkit of the applied statistician. The FRW or Bayesian bootstrap is also known as: the random-weight bootstrap, the weighted likelihood bootstrap, the weighted bootstrap, and the perturbation bootstrap (e.g., Rubin 1981, Newton and Raftery 1994, Jin, Ying, and Wei 2001).

The FRW bootstrap was first suggested by Rubin (1981), who 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) generalize 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. 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.

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 contin-

uous distribution that has a mean and standard deviation being equal (e.g., an exponential distribution with mean one). 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 FRW bootstrap methods and gives some theoretical properties of the FRW bootstrap method. Section 3 provides applications of the FRW bootstrap in CI constructions using heavily censored field-failure data, prediction intervals using data with complicated censoring, finding an appropriate model for a designed experiment, and logistic regression with rare events. Section 4 shows the results of a simulation study for bootstrap success probability and coverage probability. Section 5 provides some concluding remarks and areas for further research. Technical proofs and additional example details are given in the online supplementary material. All the computing codes are also included in the online supplementary material.

2 Integer and Fractional-Random-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 $(w_1, \ldots, w_n)'$ follow a multinomial distribution with equal event probability $1/n$.

As an illustration, the first column of Table 1 gives tree volume for 15 loblolly pine trees in units of cubic meters. The data are a subsample of the 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 ($j = 1, 2, 3$). 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

Table 1: Three integer-weight and FRW 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

resample to obtain the B bootstrap estimates.

2.2 Fractional-Random-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 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

$$f(w_1, \ldots, w_n; \alpha_1, \ldots, \alpha_n) = \frac{1}{\mathrm{B}(\alpha_1, \ldots, \alpha_n)} \prod_{i=1}^{n} w_i^{\alpha_i - 1}, \qquad \sum_{i=1}^{n} w_i = 1, \qquad w_i \geq 0, \tag{2}$$

and $\mathrm{B}(\alpha_1, \ldots, \alpha_n)$ is the normalizing factor. The uniform Dirichlet distribution is a special case where $\alpha_i = 1, i = 1, \ldots, n$. The continuous weights, like the integer multinomial resampling weights, sum to n, and have expectation 1 and 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.

Although FRW bootstrap methods were developed within a nonparametric Bayesian framework, they also apply to non-Bayesian and parametric inference problems. There are statis-

tically valid alternative methods to generate the random fractional weights (e.g., Jin, Ying, and Wei 2001). 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 parametric or semi-parametric 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 dataset will cause estimation problems, the resampling bootstrap approach will often give poor results or fail altogether. For example, in regression where one of the predictors is a factor variable with a rare level, excluding those observations makes a parameter non-estimable. Generally, when using the FRW bootstrap, because all of the original observations remain in the sample, estimation difficulties do not arise.

2.3 Bias-corrected Confidence Interval

In this section we describe the procedure to obtain a bias-corrected percentile bootstrap confidence interval (BCCI) with bootstrap estimates. Let θ be the parameter of interest and $\widehat{\theta}$ be the estimate of θ. Let $\widehat{\theta}^{(1)}, \widehat{\theta}^{(2)}, \cdots, \widehat{\theta}^{(B)}$ be the sorted bootstrap estimates in an increasing order, where B is a large number (i.e., $B = 2000$). The approximate $100(1-\alpha)\%$ BCCI for θ is

$$\left[\widehat{\theta}^{(l)}, \quad \widehat{\theta}^{(u)}\right],$$

where $l = \text{Rnd}(B\Phi_{\text{norm}}(2z_q + z_{\alpha/2}))$ and $u = \text{Rnd}(B\Phi_{\text{norm}}(2z_q + z_{1-\alpha/2}))$. Here, $z_p = \Phi^{-1}_{\text{norm}}(p)$ is the p quantile of the standard normal distribution, $\Phi_{\text{norm}}(\cdot)$ is the cumulative distribution function (cdf) of standard normal distribution, q is the proportion of the B bootstraps estimates of θ that are less than $\widehat{\theta}$, and $\text{Rnd}(\,\cdot\,)$ rounds to the nearest integer.

2.4 Theoretical Results

In this section, we present some new theoretical results that are specific to the likelihood inference for lifetime data with censoring, which provide the basis for the statistical inference for lifetime data using the FRW bootstrap.

For likelihood-based inference, 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 iid exponential distribution with mean one. Then the random vector

$$\left(\frac{Z_1}{\sum_{i=1}^n Z_i}, \ldots, \frac{Z_i}{\sum_{i=1}^n Z_i}, \ldots, \frac{Z_n}{\sum_{i=1}^n Z_i}\right)' \tag{3}$$

has a uniform Dirichlet distribution. For convenience we will use the weights from the exponential distribution with mean one going forward.

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 averaged loglikelihood function is

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

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 (3) 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 following three results give some properties of $\widehat{\boldsymbol{\theta}}^*$ and their proofs are given in the online supplement.

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 (page 136 of Shao 2003). The asymptotic normality is related to the distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}^* - \widehat{\boldsymbol{\theta}})|\boldsymbol{X}_n$, which is also a function of $\boldsymbol{X}_n$.

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

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

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 re-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 generated from commonly used log-location-scale family of distributions (e.g., the lognormal and 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.

3 Applications

3.1 Applications to Confidence Intervals

In this section, we use three examples to illustrate the applications of FRW in the construction of CIs for parameters. We present the details here for the analysis of the Bearing Cage field failure data and briefly mention the analyses of the ball bearing failure time data and rocker motor field failure data.

3.1.1 Background of Bearing Cage Field Failure Data

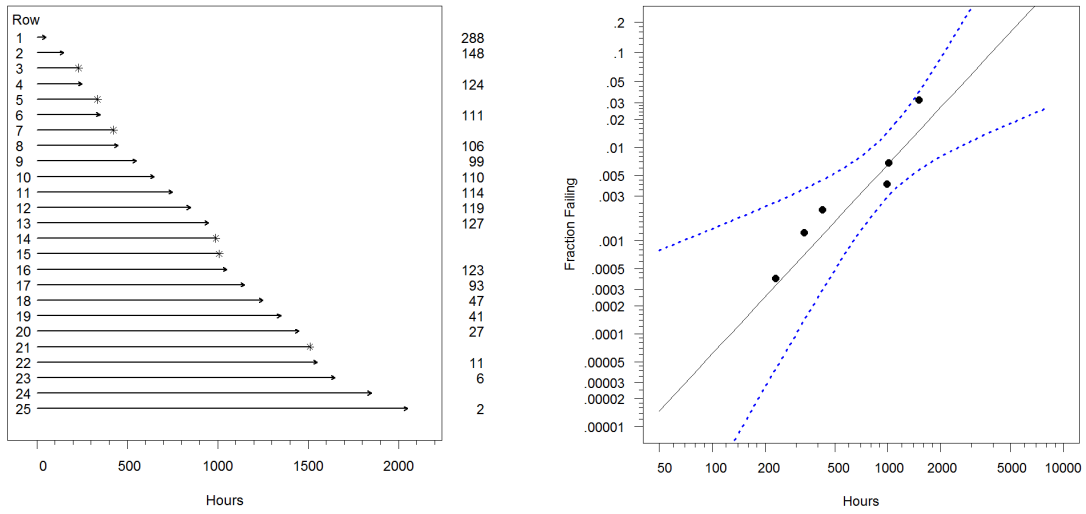
The data consist of 1703 aircraft engines put into service over time, as shown in the event plot in Figure 1(a). There were 6 failures and 1697 right-censored observations. These data were originally given in Abernethy et al. (1983) and were re-analyzed in Chapter 8 of Meeker and Escobar (1998).

3.1.2 Weibull Analysis

We use the Weibull distribution with cdf

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

as a parametric model for 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 2.03. The upper endpoint of the Wald 95% CI is 5.67 and the likelihood upper endpoint is 3.58. Another alternative for computing CIs is the bootstrap. Care is needed, however, when using the resampling bootstrap method with heavy censoring. If the expected number failing is



(a) Event Plot

(b) Probability Plot

Figure 1: Event plot and Weibull probability plot for the bearing cage field-failure data. The numbers on right column of (a) show the number of censored units at the same time.

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

Parameter	Estimate	Std Error	95% Wald CI	
			Lower	Upper
η	11792.17	9848.12	2294.67	60599.21
β	2.03	0.66	1.24	5.67

too small there could be bootstrap samples with only 0 or 1 failures, possibly causing the ML algorithm to fail. In this case, some software such as JMP-PRO will assign a large value to the estimate (e.g., 10,000). As described in **Result 3** of Section 2.4, 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 zero.

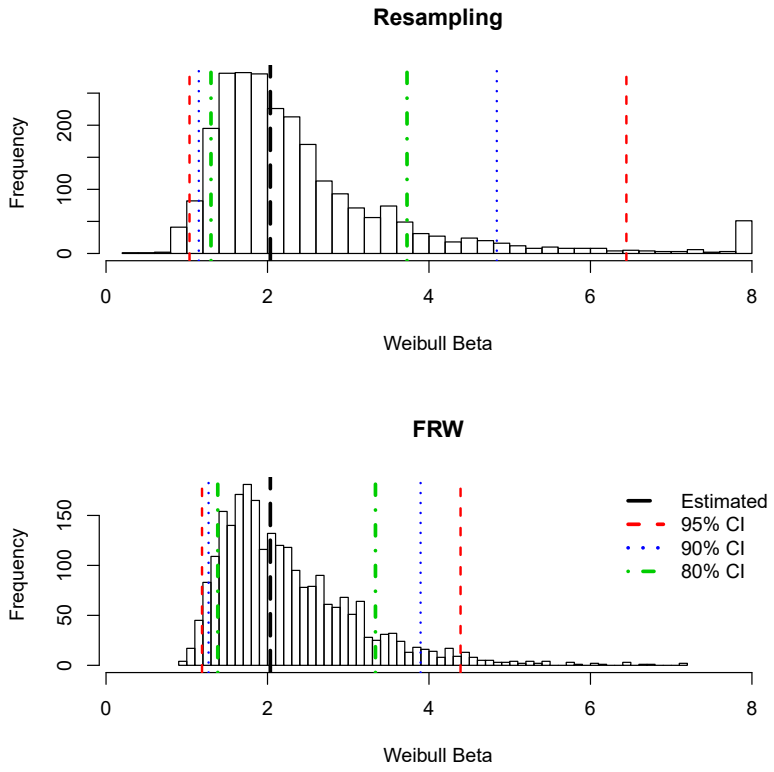


Figure 2: Histograms of the resampling and FRW bootstrap estimates with BCCIs for the Weibull shape parameter for the bearing cage field-failure data. Note that large values are truncated at 8 for better visualization.

3.1.3 Bootstrap Results

Figure 2 shows results from the resampling and the FRW bootstrap for the Weibull shape parameter β. Table 3 gives BCCIs for β. The histogram on the top shows that there were 36 samples that resulted in a wild estimate of β which were caused by having resamples with 0 failures. The upper endpoint of the 95% BCCI is larger than that provided by the Wald method. The histogram on the bottom, based on the FRW bootstrap method is better behaved and the upper endpoint is 4.4. This 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 which require extra time trying to find a maximum that does not exist.

Table 3: Resampling and FRW bootstrap results for the Weibull shape parameter for the bearing cage field-failure data.

Resampling Bootstrap Confidence Limits			FRW Bootstrap Confidence Limits		
Confidence Level	BC Lower	BC Upper	Confidence Level	BC Lower	BC Upper
0.95	1.04	6.54	0.95	1.19	4.40
0.90	1.15	4.87	0.90	1.27	3.90
0.80	1.30	3.75	0.80	1.38	3.34
0.50	1.57	2.67	0.50	1.63	2.64

3.1.4 Other Challenging Applications

Here we describe two more applications on the construction of CIs using FRW. The details of these two examples are available in Section 1 of the Supplement.

The failure analysis of the rocker motor field-failure data is particularly challenging due to heavy censoring in the data. The data first appeared in Olwell and Sorell (2001) and were reanalyzed in Chapters 14 and 18 of Meeker, Hahn, and Escobar (2017). The data consist of 1,940 rockets put into service over a period of 18 years. Among those, 1,937 of these motors performed satisfactorily (1,937 right-censored observations). There were three catastrophic launch failures but the exact failure times were unknown (yielding 3 left-censored observations). Due to heavy censoring, the resampling bootstrap method does not work properly but the FRW can be used without estimability issues in generating bootstrap estimates.

For another example, 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). 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 Fréchet ($\lambda = -1$) distributions are special cases. Thus, we are interested in the construction of a CI for λ. When the sample size is not large, the λ parameter in the generalized gamma distribution can be difficult to estimate. With the resampling method, there were ML estimate convergence problems with a substantial number of the bootstrap samples. The FRW method performed much better and provides a feasible method for computing CIs for λ.

3.2 An Application to Prediction Intervals

In this section, we illustrate the use of FRW in the construction of prediction intervals.

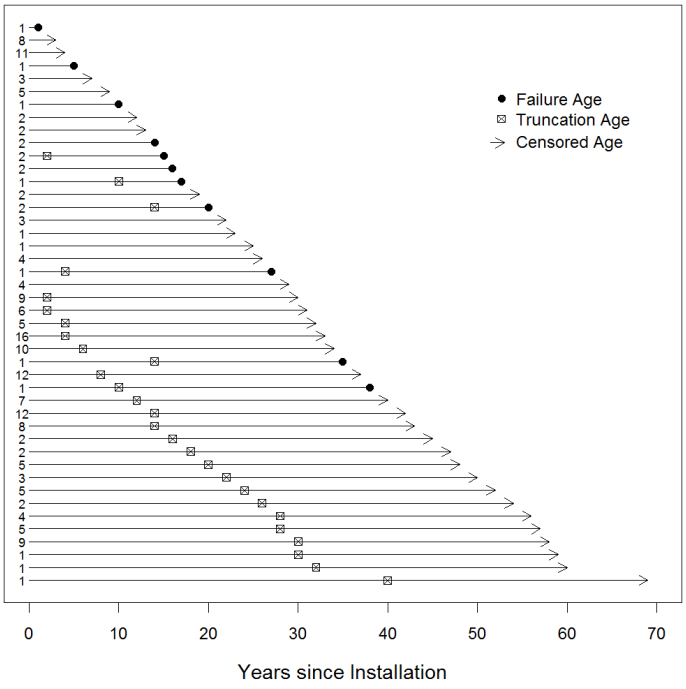


Figure 3: Event plot for the power transformer field-failure data.

3.2.1 Background

Extending the 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 having 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 installed before 1980 which were still in service are observations from a truncated distribution. Figure 3 is an event plot of a representative subset of the data.

There are several categorical covariates, including manufacturer and cooling method, that have an effect on the life distribution. Even after adjustment for the other covariates, 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 the fact that those transformers were designed to be more robust.

3.2.2 Modeling and Maximum Likelihood Estimation

We do stratification based on whether units were manufactured before or after 1987 and fit separate Weibull models to each stratum. The likelihood function is $L(\boldsymbol{\theta}|\text{DATA}) = \prod_{i=1}^{n} L_i(\boldsymbol{\theta})$, where

$$L_i(\boldsymbol{\theta}) = f(t_i;\boldsymbol{\theta})^{\delta_i\nu_i} \cdot \left[\frac{f(t_i;\boldsymbol{\theta})}{1-F(\tau_i^L;\boldsymbol{\theta})}\right]^{\delta_i(1-\nu_i)} \cdot [1-F(t_i;\boldsymbol{\theta})]^{(1-\delta_i)\nu_i} \cdot \left[\frac{1-F(t_i;\boldsymbol{\theta})}{1-F(\tau_i^L;\boldsymbol{\theta})}\right]^{(1-\delta_i)(1-\nu_i)}.$$

Here 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 $F(t;\boldsymbol{\theta})$ the pdf and cdf of the Weibull distribution, respectively. The weighted log-likelihood function can be constructed as $\sum_{i=1}^{n} w_i \log[L_i(\boldsymbol{\theta})]$.

A general prediction problem can be described as follows. Suppose one wants to predict a random quantity Y. One can determine endpoints (L, U) with the probability that Y will fall within L and U with probability $1-\alpha$. The simple "plug-in" prediction interval is obtained by taking the lower and upper quantiles of the estimated distribution of Y, which generally have poor coverage probabilities (page 294 of Meeker and Escobar 1998). Thus calibration of the simple plug-in prediction interval is needed. The basic idea of calibration is to find a nominal coverage probability $1-\alpha_c$ such that the actual coverage probability is $1-\alpha$. The bootstrap estimates are used to estimate the actual coverage probability. More details are available in Hong, Meeker, and McCalley (2009).

For the transformer application, Hong, Meeker, and McCalley (2009) used B sets of bootstrap estimates to calibrate the plug-in intervals. An important question was how to generate bootstrap samples to do the calibration. 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 (i.e., manufacturer and cooling method) and the small number of failures in some of the categories. The FRW bootstrap offered an attractive, easy-to-implement alternative.

3.2.3 Prediction Results

Figure 4 shows point predictions (i.e., the estimated mean number of failures) and 90% prediction intervals for the cumulative number of transformer failures for the next ten years, starting in 2008. The 90% confidence level is often used in prediction setting because 95% level prediction intervals tend to be wide. The prediction is made for transformers that were installed before 1987 and the number of units in the risk set is 449. For a subset of the transformers that were still in operation at the time the predictions were made, Figure 5 shows the age of the transformer and a prediction interval quantifying the information available about

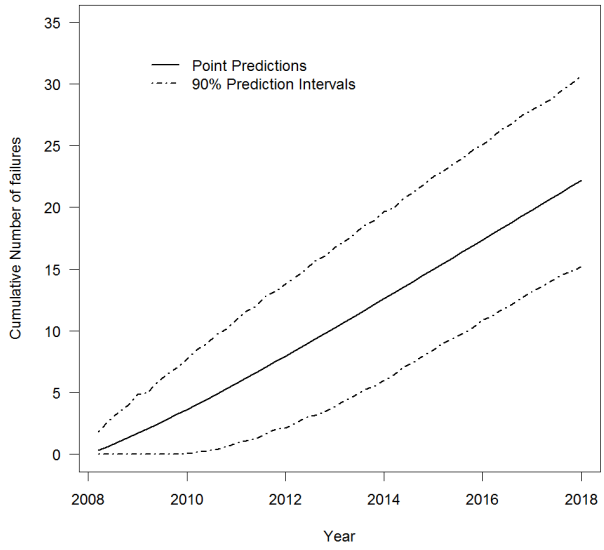


Figure 4: Power transformer fleet predictions based on the FRW bootstrap. The number of units in the risk set is 449.

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 allow a useful ranking of which transformers were at highest risk for failure in the short term.

3.3 An Application in Design of Experiments

3.3.1 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 often lack substantial redundancy, so that removing 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 (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.

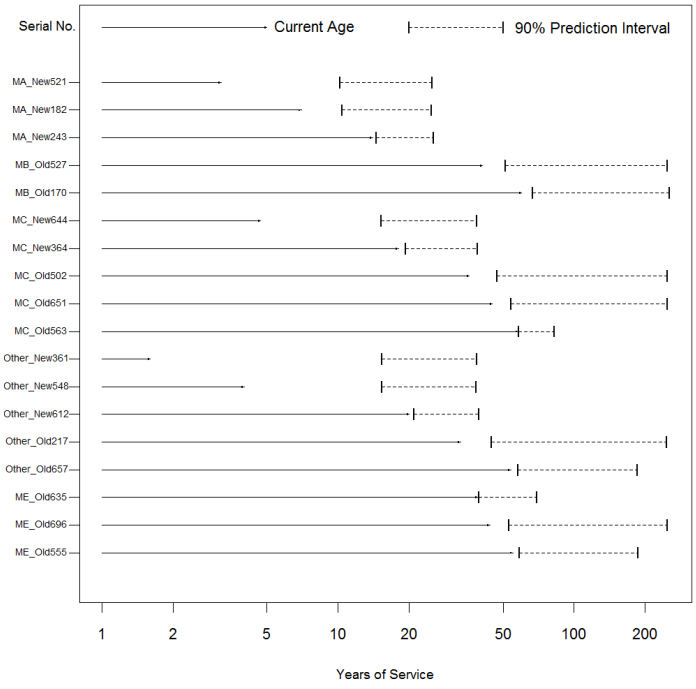


Figure 5: Power transformer individual predictions based on the FRW bootstrap.

3.3.2 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 the 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. The FRW bootstrap can keep all observations during the modeling process, and is thus suitable for model-building applications with data from a designed experiment.

3.3.3 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.e., $n = 32$) I-Optimal response surface model design was created with 7 continuous factors: Hydrogen Fraction in primary fuel, Air/Fuel Ratio, Lance Position X, Lance Position Y, Secondary Fuel Fraction, Dispersant, and Ethanol Percentage in primary fuel. This design would allow estimation of all main effects, two-factor interactions,

Table 4: 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	30.31	0.43	4939.20	<0.0001	29.46	31.150
Hydrogen Fraction	2.45	0.34	51.72	<0.0001	1.79	3.12
Air/Fuel Ratio	−2.30	0.34	44.86	<0.0001	−2.98	−1.63
Lance Position X	0.85	0.31	7.80	<0.01	0.25	1.45
Lance Position Y	0	0	0	1	0	0
Sec. Fuel Fraction	−1.10	0.26	17.82	<0.0001	−1.61	−0.59
Dispersant	0	0	0	1	0	0
Ethanol	0	0	0	1.00	0	0
Hydrogen ∗ Hydrogen	0	0	0	1	0	0

and quadratic effects. We want to assess the importance of the input variables (including the two-factor interactions and quadratic terms).

3.3.4 Using Forward Selection

First, we apply a forward stepwise procedure that selects a model using the AIC criterion. Because the computing of the effective degrees of freedom in the bootstrap samples is complicated, we use the sample size n as the degrees of freedom. The results are shown in Table 4. 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.

3.3.5 Bootstrapping the Forward Selection Procedure

We use the FRW approach to bootstrap the forward selection procedure. One thousand FRW bootstrap datasets were generated. For each FRW bootstrap dataset, 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.

Supplementary Table 6 shows partial results for the first 16 FRW bootstrap datasets (i.e., for selected regression coefficients). Figure 6 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 (e.g., frequently for Lance

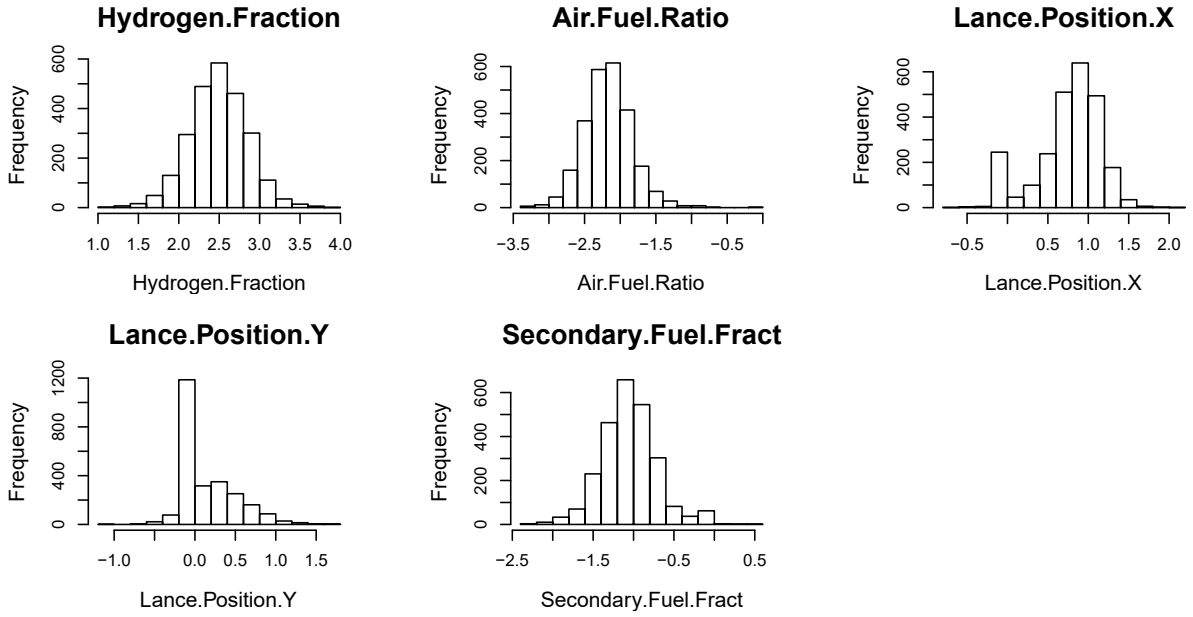


Figure 6: Histograms that summarize the FRW Bootstrap modeling results.

Position Y and never for Hydrogen Fraction). Table 5 gives the proportion of times across the 1000 bootstrap samples that each variable was chosen to be in the model. One could then use a cutoff point (such as 0.50) to decide whether or not to include model terms.

3.4 Logistic Regression for Rare Events

3.4.1 Background

In this section, we illustrate the use of the FRW bootstrap in logistic regression for rare events. Yuan et al. (2018) use survival analysis techniques to predict the time to default for companies. For illustration, we use a subset of the data from Yuan et al. (2018) and model the probability that a company will default in a specific period of time after the financial crisis in 2008 (i.e., the response variable is binary). In our dataset, we include 5,509 companies that were still in business at the start of August 2007. Of these companies, 49 defaulted in the time period between August 2007 and March 2008. The overall default rate is less than 1%. The continuous explanatory variables are Distance to Default (DTD, a widely used market-based measure of corporate default risk) and Trailing Return (returns for past specific periods). The nominal explanatory variable is company category with eight levels: Construction (1/80), Finance (2/964), Manufacturing (17/2353), Mining (2/232), Retail Trade (12/349), Services (10/859), Transportation (3/423), Wholesale Trade (2/249). The numbers in parenthesis show the proportion of defaults within each category.

Table 5: 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	1.00
Air/Fuel Ratio	1.00
Secondary Fuel Fraction	0.98
Air/Fuel Ratio $*$ Air/Fuel Ratio	0.95
Lance Position X	0.90
Lance Position X $*$ Secondary Fuel Fraction	0.85
Hydrogen Fraction $*$ Secondary Fuel Fraction	0.73
Lance Position	0.59
Dispersant	0.55
Secondary Fuel Fraction $*$ Secondary Fuel Fraction	0.48
Lance Position Y $*$ Lance Position Y	0.29
Hydrogen Fraction $*$ Lance Position Y	0.22
Hydrogen Fraction $*$ Hydrogen Fraction	0.20
Lance Position Y $*$ Dispersant	0.19
Hydrogen Fraction $*$ Lance Position X	0.18
Air/Fuel Ratio $*$ Lance Position Y	0.18
Air/Fuel Ratio $*$ Dispersant	0.16
Ethanol	0.11
Hydrogen Fraction $*$ Air/Fuel Ratio	0.10
Lance Position X $*$ Lance Position X	0.08

Table 6: ML estimates for the logistic regression analysis of the default data.

Parameter	Estimate	Std Error	95% Wald CI	
			Lower	Upper
Intercept	−3.96	0.33	−4.70	−3.38
Trailing Return	−0.25	0.56	−1.50	0.69
Distance to Default	−1.43	0.19	−1.83	−1.05
Category – Construction	−0.32	0.95	−2.90	1.18
Category – Finance	−1.44	0.66	−3.06	−0.33
Category – Manufacturing	0.35	0.32	−0.27	0.99
Category – Mining	−0.07	0.68	−1.71	1.09
Category – Retail Trade	1.33	0.36	0.61	2.04
Category – Services	0.24	0.36	−0.49	0.96
Category – Transportation	0.35	0.57	−0.95	1.37

3.4.2 Logistic Regression and Bootstrap Results

We fit a logistic regression model to describe the default outcome using the DTD, Trailing Return, and company category as explanatory variables. The Wholesale Trade level is treated as the baseline. Table 6 shows the ML estimates for the logistic regression analysis of the default data. The DTD, category-finance, and category-retail trade are statistically at the 95% level according to the Wald CI.

We compare the resampling and FRW bootstrap results for the regression coefficients for DTD and category-construction to illustrate the benefit of FRW CI construction with respect to estimating the probability of rare events. Figure 7 shows histograms of the resampling and FRW bootstrap estimates with BCCIs for the regression coefficient of DTD. Figure 8 shows similar results for the coefficient of category-construction. The detailed numbers are given in Tables 7 and 8, respectively. The histograms and BCCIs of other covariates are available in Supplementary Section 4.

For the continuous variables in the logistic regression, the resampling and FRW bootstrap behave similarly. However, for the categorical covariate, the FRW bootstrap estimates are much more stable. Resampling bootstrap produces many extremely small estimates compared to the result estimated with the original data. The reason is that, in each resampling, it is possible that the bootstrap sample has no default event within the company category-construction, because a default is a rare event and less than 1% of total companies defaulted. Thus, the resampling procedure can suffer from an estimability issue, which can be easily avoided by using the FRW bootstrap procedure.

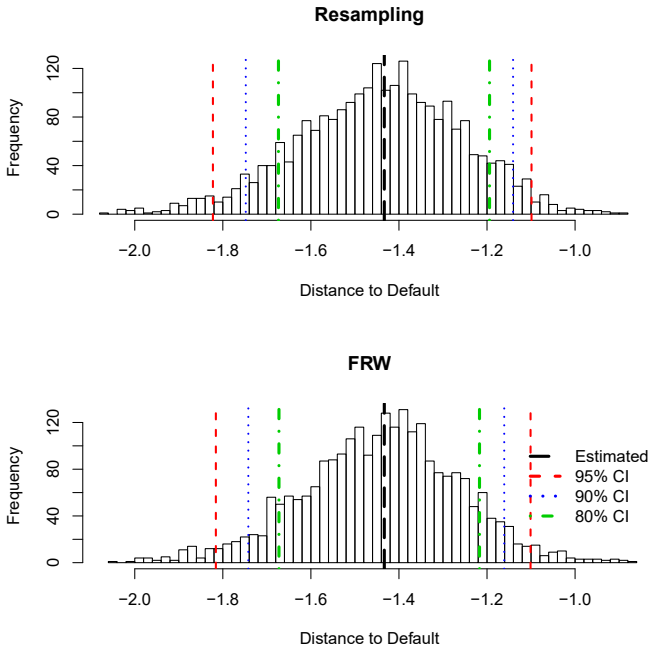


Figure 7: Histograms of resampling and FRW bootstrap estimates with BCCIs for the regression coefficient of DTD for the default data.

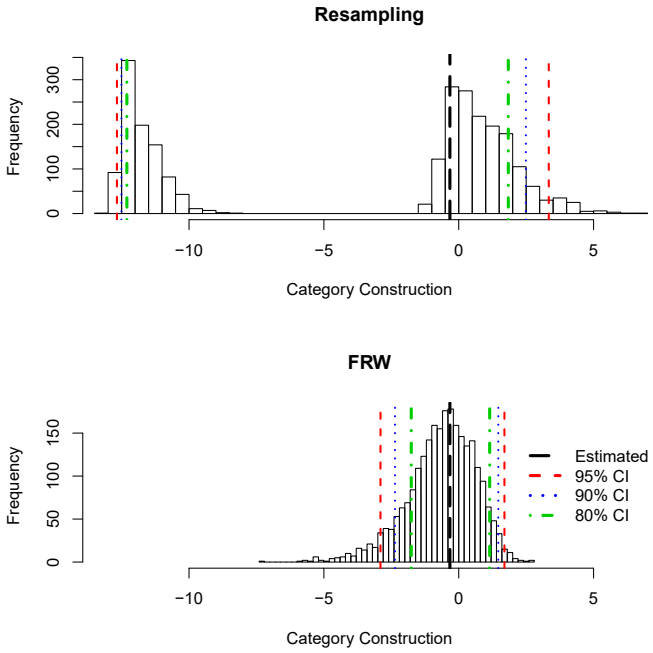


Figure 8: Histograms of resampling and FRW bootstrap estimates with BCCIs for the regression coefficient of category-construction for the default data.

Table 7: Resampling and FRW bootstrap results for the regression coefficient of DTD for the default data.

Resampling Bootstrap Confidence Limits			FRW Bootstrap Confidence Limits		
Coverage	BC Lower	BC Upper	Coverage	BC Lower	BC Upper
0.95	−1.82	−1.09	0.95	−1.81	−1.09
0.90	−1.74	−1.14	0.90	−1.73	−1.16
0.80	−1.67	−1.19	0.80	−1.67	−1.21
0.50	−1.55	−1.30	0.50	−1.55	−1.32

Table 8: Resampling and FRW bootstrap results for the regression coefficient of category-construction for the default data.

Resampling Bootstrap Confidence Limits			FRW Bootstrap Confidence Limits		
Coverage	BC Lower	BC Upper	Coverage	BC Lower	BC Upper
0.95	−12.67	3.33	0.95	−2.90	1.69
0.90	−12.50	2.49	0.90	−2.36	1.46
0.80	−12.30	1.83	0.80	−1.76	1.14
0.50	−11.85	0.73	0.50	−0.91	0.60

4 Simulation Study

In this section, we conduct a small scale simulation study to compare the estimability and CI coverage property of the resampling and FRW bootstrap methods.

4.1 Simulation Design

We use a simulation setting that is similar to Jeng and Meeker (1999). The data are simulated from the Weibull distribution with time (Type I) censoring. In the study, we also consider the following two factors: p_f, the probability of failure of each sample, and $\mathrm{E}(r)$, the expected number of failures in each sample. The values for p_f and $\mathrm{E}(r)$ are given in Table 9. In total, there are $8 \times 11 = 88$ combinations. Following Jeng and Meeker (1999), we only keep those datasets that can generate ML estimates and then we do the bootstrap.

We first investigate the estimability of parameters based on bootstrap samples. The success proportion provides an estimate for the probability that each single bootstrap sample can estimate the Weibull parameters. From the theoretical results in Section 2.4, the FRW method can guarantee estimability as long as the original dataset can generate ML estimates. Because we only do the bootstrap for those datasets that can generate ML estimates, the FRW bootstrap method will always generate ML estimates in our simulation study. For the resampling method, it is possible that re-sampled data do not result in ML estimates. That is, the bootstrap sample will not succeed in estimating parameters, causing an estimability problem.

We also investigate the coverage probability (CP) of CIs constructed by using estimates from both bootstrap methods. Both the resampling and FRW bootstrap methods are used to construct two-sided CI and one-sided confidence bounds for the Weibull parameters β and η. In the paper, we present the CP results for one-sided confidence bounds, as one can deduce the two-sided CP from those results. The results for two-sided CIs are available on the online supplement.

We simulated 10000 datasets for each p_f and $\mathrm{E}(r)$ combination. For each dataset, we do resampling and FRW bootstrap 2000 times. The BCCI and one-sided confidence bounds are computed. The CP for one simulation setting is the proportion of trials for which the BCCI contains the true parameter. We also record the number of bootstrap samples for which ML estimates could not be computed. Those bootstrap samples that can not generate ML estimates are excluded from the CI computation.

Table 9: The values for parameters p_f and $\mathrm{E}(r)$.

Parameter	Levels
p_f	$0.01, 0.03, 0.05, 0.1, 0.2, 0.5, 0.7, 0.9$
$\mathrm{E}(r)$	$5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100$

4.2 Results and Discussions

Figure 9 is a heat map for the success probability for the resampling bootstrap for various combinations of p_f and $\mathrm{E}(r)$. From the figure, we can see that the resampling bootstrap can have a significant number of bootstraps fail when $\mathrm{E}(r)$ is small. The FRW bootstrap, however, is robust and does not have an estimation problem (i.e., the probability is 1 for all combinations). From this result, we observe that the resampling bootstrap is likely to fail when the sample size is small, resulting in estimability and interpretation problems. The FRW does not suffer these estimability problems, makes it more useful in certain practical applications (as illustrated in our numerical examples in Section 3).

Figure 10 plots the CP versus $\mathrm{E}(r)$ for the Weibull β parameter using resampling and FRW bootstrap estimates for various p_f. The left and right panels show the results for the 95% one-sided lower and upper confidence bounds, respectively. The results for 90% BCCI for β, and BCCI and one-sided confidence bounds for η are available in Section 5 of the Supplement. For the one-sided lower confidence bound, both methods have CP close to the nominal 95%, even when $\mathrm{E}(r)$ is around 5. The CP for both methods, however, are quite close to each other, though the FRW method has slightly better results. For the one-sided upper confidence bound, both bootstrap methods have CP close to 95% when $\mathrm{E}(r)$ is greater than 20. When $\mathrm{E}(r) \leq 10$, both methods have CP values that are smaller than the nominal level 0.95. The p_f value does not have significant effect on the CP. Comparing the two bootstrap methods, the difference in CP is small, especially when $\mathrm{E}(r)$ is large.

Overall, the FRW method outperforms the resampling method in term of estimability, while both have comparable performance in terms of CP. We note that the CP tends to be small, which is always a challenging problem when one needs to deal with small samples. Small sample techniques such as generalized pivotal quantities (e.g., Chapter 14 of Meeker, Hahn, and Escobar 2017), the Bartlett-corrected likelihood procedure, and bootstrap-calibrated likelihood-ratio procedures (e.g., Jeng and Meeker 1999) can be used to improve the CP when $\mathrm{E}(r)$ is small.

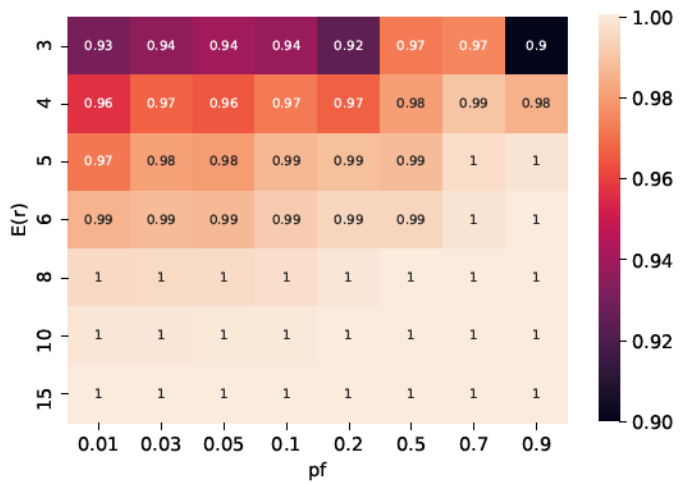
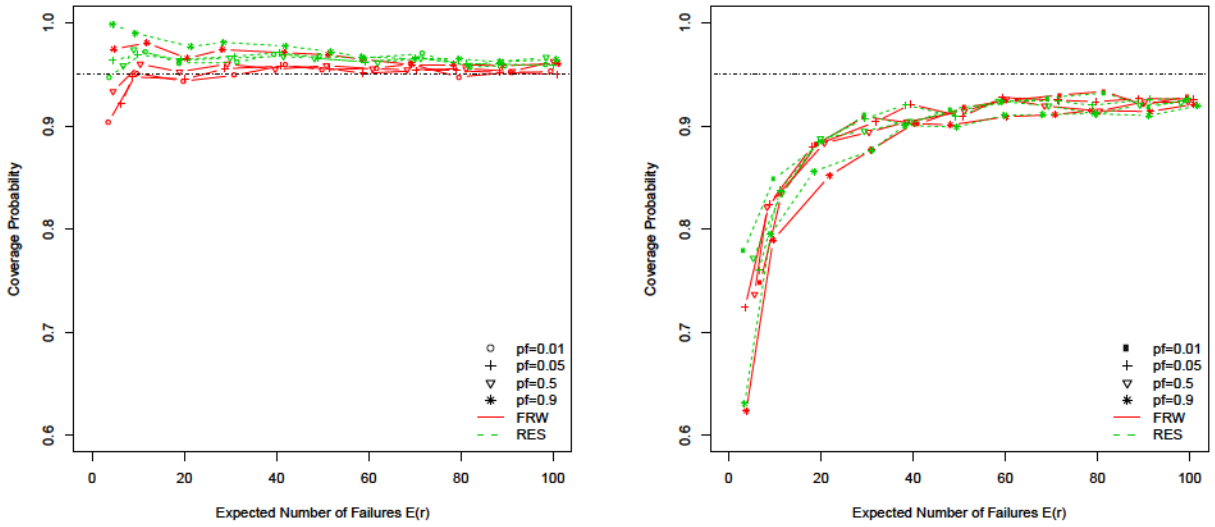


Figure 9: Heat map for the probability of success estimation for the resampling bootstrap under various combinations of p_f and E(r).



(a) 95% Lower Confidence Bound

(b) 95% Upper Confidence Bound

Figure 10: Plot of CP versus E(r) for the One-Sided confidence bounds for the Weibull β parameter using resampling and FRW bootstrap estimates for various p_f. Note that the x-locations of the points are jittered and only a subset of p_f values are plotted for better visualization.

5 Conclusions and Areas for Future Research

5.1 Concluding Remarks

With vastly improved computing capabilities and bootstrap theory that has been developed over the past 40 years, bootstrapping provides an important useful tool for obtaining CIs, prediction intervals, and better regression models. The FRW bootstrap tremendously expands the potential areas of application of the bootstrap to applications involving heavy censoring and/or truncation, categorical explanatory variables, and designed experiments where dropping certain combinations of the original observations can cause estimability problems. As illustrated in our examples, the FRW bootstrap has far fewer problems with estimability than the resampling bootstrap when dealing with censoring and/or truncation and categorical explanatory variables, which are common in practical applications.

Overall, we observe that the FRW bootstrap is as easy to implement as the resampling bootstrap and it has similar desirable properties 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 have sought to present the FRW bootstrap as a safer, more broadly applicable, alternative to the resampling bootstrap.

The software we use for analysis in this paper is JMP-PRO and R. We use JMP-PRO for survival analysis for the real datasets in Section 3. For the simulation study in Section 4, we used R to do the bootstrap and obtain the ML estimates.

5.2 Areas for Future Research

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. There are different, asymptotically equivalent ways to choose the random weights 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?

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 it is 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. It also would be useful to have

FRW analogs of second-order correct CIs such as the bias-corrected and accelerated (BCa) and bootstrap-t methods.

Generalized fiducial inference (GFI) has proven to be a powerful tool for defining CI procedures for non-standard models (see Hannig, Iyer, and Patterson 2006, Majumder and Hannig 2016, and Hannig et al. 2016). Implementing GFI methods generally requires computing a large set of simulated parameter estimates, in a manner 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 a wider range of applications.

Supplementary Materials

The following supplementary material is available online.

Additional details: Technical details, additional results for applications, and graphs on additional simulation results (pdf file).

Code and data: Datasets, JMP and R code for simulations and data analysis. (zip file).

Acknowledgments

The authors thank Dan Nordman from Iowa State University for his helpful comments on an earlier version of the paper. The authors also thank the editor, an associate editor, and two referees for their valuable comments that helped in improving this paper significantly. The authors acknowledge Advanced Research Computing at Virginia Tech for providing computational resources. The research by Hong and Xu was partially supported by National Science Foundation Grants CNS-1838271 and CMMI-1904165 to Virginia Tech.

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. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*, Volume 38. SIAM.

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.

Jeng, S.-L. and W. Q. Meeker (1999). Comparisons of approximate confidence interval procedures for type I censored data. *Technometrics 42*, 135–148.

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.

Lo, A. Y. (1993, 03). A Bayesian bootstrap for censored data. *The Annals of Statistics 21*(1), 100–123.

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. (2003). *Mathematical Statistics*. New York, NY: Springer-Verlag.

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.

Yuan, M., C. Tang, Y. Hong, and J. Yang (2018). Disentangling and assessing uncertainties in multiperiod corporate default risk predictions. *The Annals of Applied Statistics 12*, 2587–2617.