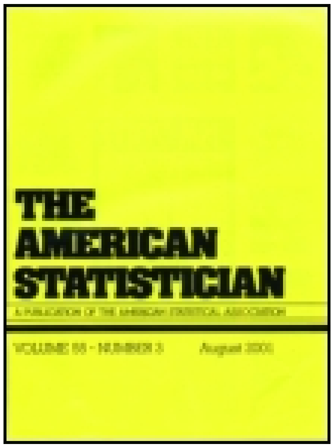

The American Statistician

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

Applying Bootstrap Methods to System Reliability

Christopher E. Marks, Andrew G. Glen, Matthew W. Robinson & Lawrence M. Leemis

Accepted author version posted online: 11 Jun 2014.Published online: 11 Aug 2014.

To cite this article: Christopher E. Marks, Andrew G. Glen, Matthew W. Robinson & Lawrence M. Leemis (2014) Applying Bootstrap Methods to System Reliability, The American Statistician, 68:3, 174-182, DOI: 10.1080/00031305.2014.928232

To link to this article: http://dx.doi.org/10.1080/00031305.2014.928232

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TAS

Applying Bootstrap Methods to System Reliability

Christopher E. MARKS, Andrew G. GLEN, Matthew W. ROBINSON, and Lawrence M. LEEMIS

We present a fully enumerated bootstrap method to find the empirical system lifetime distribution for a coherent system modeled by a reliability block diagram. Given failure data for individual components of a coherent system, the bootstrap empirical system lifetime distribution derived here will be free of resampling error. We further derive distribution-free expressions for the bias associated with the bootstrap method for estimating the mean system lifetimes of parallel and series systems with statistically identical components. We show that bootstrapping underestimates the mean system lifetime for parallel systems and overestimates the mean system lifetime for series systems, although both bootstrap estimates are asymptotically unbiased. The expressions for the bias are evaluated for several popular parametric lifetime distributions. Supplementary materials for this article are available online.

KEY WORDS: Bias; Bootstrapping; Parametric lifetime distributions.

1. INTRODUCTION

Bootstrapping as a statistical method is ubiquitous. In most applications, one is left with two types of error: random sampling error associated with the dataset and resampling error associated with the bootstrapping process. Resampling error is generally contained, but not eliminated, by increasing the number of bootstrap iterations, B. The "ideal bootstrap" sets $B = \infty$ (Efron and Tibshirani 1993), and is sometimes calculable by fully enumerating all possible outcomes of a bootstrap instead of resampling the data. In this article, we present a fully enumerated bootstrap method that has this advantage of eliminating resampling error. We will apply the method to models of a coherent system, represented by a reliability block diagram (RBD). This fully enumerated bootstrap produces the distribution function of a discrete random variable that approximates the system's continuous lifetime distribution. We show that there is inherent bias that can be computed exactly under various parametric assumptions. System designers can use these methods to investigate system parameters to include in the design, for example, the effect of changing the number of components in series, parallel or combinations of both to achieve system improvement.

The article is organized as follows. In Section 2, we investigate the simplest forms of RBDs consisting of two components whose lifetimes are independent and identically distributed (IID) random variables and the components are placed in parallel (or series) to illustrate the nature of the underlying mathematics in the enumerated bootstrap. In Section 3 we present the steps of a resampled bootstrap and present a simple example of how bias is induced by the "minimum" and "maximum" operations. In Section 4, we present the general result for bias for parallel and series systems. We also present some of the difficulties encountered in generalizing to systems consisting of more than two components. Section 5 presents some applications of a fully enumerated bootstrap for an RBD example.

The literature concerning bootstrapping of RBDs is sparse at best. Leemis (2006) used enumerated bootstrapping of binary data that represents the availability of components in an RBD to construct a lower confidence bound for three components in series. Doss and Chiang (1994) presented a traditional bootstrap analysis of RBDs to include a comparison of resampling schemes. There is ample literature on enumerated bootstraps, but little of it applies to RBD models.

Parallel systems have two or more components operating simultaneously that perform identical tasks as backups to each other. This system configuration occurs quite often in design. Series systems of IID components are less frequently observed. Individual fins in a turbine engine could be considered to comprise a series system: when the first one fails the system fails. Clearly, the independence of the fins could be questioned. Consider the RBD in Figure 1, a model of a computer server system (E-mail or brokerage servers perhaps). The series configuration could represent the major subsystems of power supply, computation hardware, and external memory. Each subsystem is made more resilient by having two components with IID lifetimes operating simultaneously in parallel. Thus, A_1 and A_2 represent the IID lifetimes of two power supplies, and the other components represent computer and memory subsystems. The system lifetime S is given by the algebraic relationship

$$S = \min\left\{\max\left\{A_1, A_2\right\}, \max\left\{B_1, B_2\right\}, \max\left\{C_1, C_2\right\}\right\}.$$

Given the hypothetical case that all lifetime cumulative distribution functions (CDFs) of the components are known to be $F_A(t)$, $F_B(t)$, and $F_C(t)$ respectively, the exact CDF of S is

$$F_S(t) = 1 - \left[1 - F_A^2(t)\right] \cdot \left[1 - F_B^2(t)\right] \cdot \left[1 - F_C^2(t)\right] \qquad t > 0.$$

Christopher E. Marks is a Major in the U.S. Army (E-mail cemarks@nps.edu). Andrew G. Glen is a Professor Emeritus from the United States Military Academy, West Point, NY 10996 (E-mail: Andrew.Glen.1984@gmail.com). Matthew W. Robinson is a second lieutenant in the U.S. Army (E-mail: mthw.wm.robinson@gmail.com). Lawrence M. Leemis is a Professor, Department of Mathematics, The College of William & Mary, Williamsburg, VA 23187–8795 (E-mail: leemis@math.wm.edu). The authors thank the editors and anonymous reviewers that added significantly to the presentation of this material. Further, the authors thank Lieutenant James Fritz who conducted preliminary research in the early stages of this project as a cadet at the United States Military Academy.

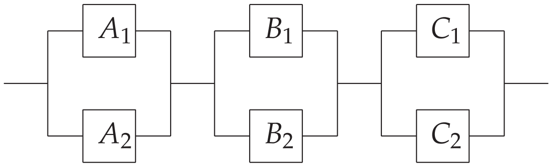

Figure 1. Block diagram of a series arrangement of parallel subsystems.

Now consider the more common case in which the CDFs of the components are not known, but life-test data are available for each component. Let n_a be the sample size of the dataset for component A, with similar definitions for n_b and n_c. These datasets, $\{a_1, a_2, \ldots, a_{n_a}\}$, $\{b_1, b_2, \ldots, b_{n_b}\}$, and $\{c_1, c_2, \ldots, c_{n_c}\}$, provide the supports for three different discrete random variables A^*, B^*, and C^*, each having equal probability for these support values (assuming that the data values are distinct). Thus, A^*, B^*, and C^* represent the random variables associated with the bootstrap samples from these three datasets. The discrete CDFs of A^*, B^*, and C^*, denoted by $F_{A^*}(t)$, $F_{B^*}(t)$, and $F_{C^*}(t)$, are the same as the empirical distribution functions (EDFs) of their corresponding lifetime datasets. Thus, a fully enumerated bootstrap of the system can be represented algebraically by

$$S^* = \min\left\{\max\{A_1^*, A_2^*\},\ \max\{B_1^*, B_2^*\},\ \max\{C_1^*, C_2^*\}\right\}.$$

The discrete CDF (that approximates the unknown, but true continuous CDF $F_S(t)$) is

$$F_{S^*}(t) = 1 - \left(1 - [F_{A^*}(t)]^2\right) \cdot \left(1 - [F_{B^*}(t)]^2\right) \cdot \left(1 - [F_{C^*}(t)]^2\right) \quad t > 0.$$

It is this CDF that is the fully enumerated bootstrapped distribution of the system. We will show that such bootstrapped systems have bias resulting from taking minimums and maximums of discrete random variables.

We use the following notation in the article:

- X — A continuous random variable with probability density function (PDF) $f_X(t)$ and CDF $F_X(t)$ that represents a component lifetime.
- $\mathbf{X}$ — A random sample of n component lifetimes.
- $X_{(i)}$ — The ith ordered element of $\mathbf{X}$, $i = 1, 2, \ldots, n$.
- S — A continuous random variable representing the system lifetime of a system consisting of two components arranged in parallel, each with an independent lifetime distributed according to X, that is, $S = \max\{X_1, X_2\}$.
- T — A continuous random variable representing the system lifetime of a system consisting of two components arranged in series, each with an independent lifetime distributed according to X, that is, $T = \min\{X_1, X_2\}$.
- X^* — A random variable distributed according to the EDF associated with $\mathbf{X}$.
- S^* — A random variable representing the system lifetime of a system consisting of two components arranged in parallel, each with an independent lifetime distributed according to X^*, that is, $S^* = \max\{X_1^*, X_2^*\}$. The distribution of S^* is used to approximate the unknown distribution of S. Also, it is known that S^* converges almost surely to S asymptotically by the Glivenko–Cantelli theorem (see, e.g., Billingsley 1995).
- T^* — A random variable representing the system lifetime of a system consisting of two components arranged in series, each with an independent lifetime distributed according to X^*, that is, $T^* = \min\{X_1^*, X_2^*\}$.

2. TWO AND *m* IID COMPONENTS IN PARALLEL OR IN SERIES

In this section, we present the mathematics underlying simpler systems, setting the foundation for the proof in the Appendix A. Consider a system with two identical components in parallel, in other words, a system with two constantly operating components with identically distributed failure times that act as real-time backups to one another. We assume that prior testing of n components has produced a failure-time dataset $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$, from which we define X^* with CDF $F_{X^*}(t) = i/n$ for $t \in [x_{(i)}, x_{(i+1)})$, $i = 1, 2, \ldots, n$ and $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ are the ordered observations. Knowing only the failure-time data (with no assumption of an underlying parametric failure-time distribution), we can approximate the true system lifetime using the random variable $S^* = \max\{X_1^*, X_2^*\}$. The CDF of S^* is

$$\begin{aligned} F_{S^*}(t) &= P(\max\{X_1^*, X_2^*\} \le t) \\ &= P(X_1^* \le t) \times P(X_2^* \le t) \\ &= F_{X^*}(t) \times F_{X^*}(t) \\ &= \frac{i^2}{n^2}, \ \text{ for } t \in [x_{(i)}, x_{(i+1)}). \end{aligned}$$

At each support point $x_{(i)}$ the value of the the probability mass function (PMF) can be found by subtracting the appropriate CDF values:

$$\begin{aligned} p_{S^*}(t) &= F_{S^*}(t) - F_{S^*}(x_{(i-1)}) \\ &= \frac{i^2 - (i-1)^2}{n^2} \\ &= \frac{2i - 1}{n^2} \text{ for } t = x_{(i)}, \quad i = 1, 2, \ldots, n. \end{aligned}$$

Note that one must assign $x_{(0)} = 0$ and $F_{S^*}(x_{(0)}) = 0$ for the indexing to be correct. We note several interesting properties of the PMF of S^*:

- The PMF of S^* has the same probability values $p_{S^*}(t)$ regardless of the values x_i in the dataset, as long as the data values are distinct (i.e., there are no repeated or tied values) for a fixed sample size n.
- The supports of S^* and X^* are identical.
- The PMF of S^* is an increasing function over the values of its support with initial mass value $p_{S^*}(x_{(1)}) = 1/n^2$, constant increments of $2/n^2$, and final mass value $p_{S^*}(x_{(n)}) = (2n-1)/n^2$ for distinct data values.

Similar methods can be used to derive the empirical system distribution S_m^*, which is the lifetime of a system of m components with IID lifetimes arranged in parallel, using X^* to approximate the lifetime distribution of each. The CDF and PMF of S_m^* are

$$F_{S_m^*}(t) = \frac{i^m}{n^m}, \quad \text{for } t \in \left[x_{(i)}, x_{(i+1)}\right),\ i = 1, 2, \ldots, n$$

and

$$p_{S_m^*}(t) = \frac{i^m}{n^m} - \frac{(i-1)^m}{n^m}, \quad \text{for } t = x_{(i)},\ i = 1, 2, \ldots, n.$$

Using a binomial expansion and simplifying, the PMF can be expressed as

$$p_{S_m^*}(t) = n^{-m} \sum_{j=1}^{m} \binom{m}{j} (-1)^{j-1} i^{m-j}, \quad \text{for } t = x_{(i)},$$
$$i = 1, 2, \ldots, n.$$

The minimum lifetime of IID components can be derived in a similar fashion. Let $T^* = \min\{X_1^*, X_2^*\}$ be the analogous discrete random variable for a system of two IID components in series, given a component failure-time dataset. The CDF and PMF of T^* are

$$F_{T^*}(t) = 1 - \left(1 - \frac{i}{n}\right)^2, \quad \text{for } t \in [x_{(i)}, x_{(i+1)}),\ i = 1, 2, \ldots, n$$

and

$$p_{T^*}(t) = \frac{2n - 2i + 1}{n^2}, \quad \text{for } t = x_{(i)},\ i = 1, 2, \ldots, n.$$

The PMF of T^* is now a decreasing function over its support, beginning with $p_{T^*}(x_{(1)}) = (2n-1)/n^2$ and ending with $p_{T^*}(x_{(n)}) = 1/n^2$, with decrements of $2/n^2$ for distinct data values. An extension for a system of m components in series can be derived.

3. BOOTSTRAPPING AND BIAS FOR A PARALLEL SYSTEM

We now illustrate how bias enters into bootstrapping. Consider the case of a two-component parallel system with IID unit exponential component lifetimes. The interest is in the mean system lifetime. The analyst is unaware of the true lifetime distribution, and only has the values of $n = 3$ component lifetimes. This small value of n is for enumeration purposes only; bootstrapping applications usually rely on larger sample sizes. What is the bias associated with the use of bootstrapping to estimate the mean system lifetime?

The unit exponential component distribution, represented by X in our notation, has expected value $E(X) = 1$ and CDF $F_X(t) = 1 - \exp(-t), \quad t > 0$. In this case, the system lifetime S has CDF

$$F_S(t) = P(S \le t) = \left(1 - \exp(-t)\right)^2 \qquad t > 0.$$

So the true expected system lifetime is

$$E(S) = \int_0^\infty \left(1 - F_S(t)\right) dt = \frac{3}{2},$$

which uses the technique for calculating the mean of a nonnegative random variable from Meeker and Escobar (1998, p. 77). The analyst, however, knows only (a) that the component lifetimes are independent, (b) that the components have identical lifetime distributions, and (c) the values of three component lifetimes: x_1, x_2, and x_3. Because there is no knowledge of the true lifetime distribution, the analyst might consider a simple resample-style bootstrapping, rather than a parametric approach or our fully enumerated approach. To further simplify the example we will set the bootstrap resample value artificially low: $B = 3$. In this setting, a resampling bootstrap proceeds as follows:

1. Three values are sampled with replacement from the dataset, representing three lifetimes from component 1, y_1, y_2, and y_3.
2. Three additional values are sampled with replacement from the dataset, representing three lifetimes from component 2, z_1, z_2, and z_3.
3. Pairwise maximums of the values from the two previous steps are taken, representing three system lifetimes, $s_1 = \max\{y_1, z_1\}$, $s_2 = \max\{y_2, z_2\}$, and $s_3 = \max\{y_3, z_3\}$.
4. The three system lifetimes, which approximate three true system lifetimes, are averaged to estimate the mean system lifetime.

To illustrate these steps with numerical values, consider the dataset of the $n = 3$ component observations 0.8, 1.7, 0.4. The steps given above might go as follows:

1. The component 1 lifetimes sampled are 0.8, 0.4, 0.8.
2. The component 2 lifetimes sampled are 1.7, 1.7, 0.4.
3. The pairwise maximums representing system lifetimes are 1.7, 1.7, 0.8.
4. The estimate for the mean system lifetime is $(1.7 + 1.7 + 0.8)/3 = 1.4$.

To calculate the bias in this estimate, consider the more general case in which the value sampled for the lifetime of a component is equally likely to be one of the data values X_1, X_2, or X_3. Therefore, there are $3 \cdot 3 = 9$ different outcomes by the multiplication rule, as illustrated below.

Component 1	$X_{(1)}$	$X_{(1)}$	$X_{(1)}$	$X_{(2)}$	$X_{(2)}$	$X_{(2)}$	$X_{(3)}$	$X_{(3)}$	$X_{(3)}$
Component 2	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$
Maximum	$X_{(1)}$	$X_{(2)}$	$X_{(3)}$	$X_{(2)}$	$X_{(2)}$	$X_{(3)}$	$X_{(3)}$	$X_{(3)}$	$X_{(3)}$

Since the nine outcomes are equally likely, the probabilities associated with the first, second, and third order statistic corresponding to the maximum of the two lifetimes in the bootstrap sample are 1/9, 3/9, and 5/9. Therefore, the expected system lifetime using bootstrapping is

$$E\left(S^*\right) = \frac{1}{9} E\left(X_{(1)}\right) + \frac{3}{9} E\left(X_{(2)}\right) + \frac{5}{9} E\left(X_{(3)}\right).$$

Using the standard order statistic results (see, e.g., Arnold et al. 2008) for random variables from a continuous population to calculate these expected values:

$$E\left(S^*\right) = \frac{1}{9} \cdot \frac{1}{3} + \frac{3}{9} \cdot \frac{5}{6} + \frac{5}{9} \cdot \frac{11}{6} = \frac{4}{3}.$$

So the bootstrapping approach underestimates the true expected system lifetime of 3/2. Specifically the bias is $E(S^*) - E(S) = 4/3 - 3/2 = -1/6$. The next section of this article will use the previous section's results to prove a general result that quantifies this bias for such parallel systems.

4. BIAS RESULT FOR TWO AND *m* IID COMPONENTS IN PARALLEL OR IN SERIES

The example in the previous section has shown that bootstrapping can underestimate the mean system lifetime of a simple parallel system. The following theorems indicate that this result is true in general and that the bootstrap estimate is asymptotically unbiased. Theorem 1 establishes that bootstrapping underestimates the mean system lifetime for a two-component parallel system. In fact, since $E(S) > E(X)$, the bias will always be negative. The result can be interpreted as follows: the bias is $-1/n$ times the expected improvement of increasing from one component to two components in parallel. Theorem 2 states that bootstrapping overestimates the mean system lifetime for a two-component series system and that the magnitudes of the two biases are equal.

Theorem 1. For a system consisting of two independent, statistically identical components with finite population means arranged in parallel, the bias resulting from the bootstrap estimation of the mean system lifetime is

$$E(S^*) - E(S) = -\frac{E(S) - E(X)}{n}.$$

Theorem 2. For a system consisting of two independent, statistically identical components with finite population means arranged in series, the bias resulting from the bootstrap estimation of the mean system lifetime is

$$E(T^*) - E(T) = \frac{E(X) - E(T)}{n} = -\big(E(S^*) - E(S)\big),$$

which, as indicated, is the negative of the result for the bootstrapping bias for a parallel system.

The proof to Theorem 1 is given in Appendix A. The proof to Theorem 2 is similar to that of the first, and is presented in this journal's online supplementary materials section. Both proofs start with the underlying properties of X^* and S^* that were outlined in Section 2. The difficult part of the proof is finding $E(S^*)$, after that the result readily follows. Table 1 contains expressions for the bias for several popular survival distributions with positive support, positive scale parameter η, and positive shape parameter κ. Note that for the gamma distribution, ${}_2F_1(\cdot)$ refers to the regularized hypergeometric function, which is reviewed with a bibliography by Weisstein (2013) and can be calculated in numerous software packages including Mathematica and Maple (see Wolfram 2011 and Maplesoft 2012). The exact form of ${}_2F_1(\cdot)$ is given in the Appendix B. The gamma distribution bias formula when $\kappa = 1$ simplifies to the corresponding exponential distribution bias formula given in the table.

Table 1. Bootstrapping bias for parametric survival distributions

Distribution	$f(t)$	$E(S^*) - E(S)$
gamma(η, κ)	$\dfrac{x^{\kappa-1}\exp(-t/\eta)}{\eta^\kappa \Gamma(\kappa)}$	$-\dfrac{\eta\Gamma(1+2\kappa)\,{}_2F_1(\kappa, 1+2\kappa, 2+\kappa, -1)}{n\Gamma(\kappa)}$
Weibull(η, κ)	$\dfrac{\kappa}{\eta}\left(\dfrac{t}{\eta}\right)^{\kappa-1}\exp\left[-\left(\dfrac{t}{\eta}\right)^{\kappa}\right]$	$-\dfrac{\eta\left(1-2^{-1/\kappa}\right)\Gamma(1+1/\kappa)}{n}$
exponential(η)	$\eta^{-1}\exp(-t/\eta)$	$-\dfrac{\eta}{2n}$
log logistic(η, κ)	$\dfrac{\kappa(t/\eta)^{\kappa-1}}{\eta[1+(t/\eta)^\kappa]^2}$	$-\dfrac{\eta\pi}{n\kappa^2\sin(\pi/\kappa)}, \quad \kappa > 1$

Some distributions do not have closed form expressions for bias, such as the log-normal distribution. Even in these cases of no closed form, numerical methods yield bias calculations. Given that $F_X(t)$ is available, the bias can be calculated analytically or numerically with the following application of Theorem 1:

$$-\frac{1}{n}\big(E(S) - E(X)\big) = -\frac{1}{n}\Big(\int_0^\infty \big(1 - F_X^2(t)\big)dt - \int_0^\infty \big(1 - F_X(t)\big)dt\Big) = -\frac{1}{n}\int_0^\infty F_X(t)\big(1 - F_X(t)\big)\,dt.$$

This result has a useful graphical interpretation. The integrand is the multiplication of the individual component's CDF times its survivor function (SF) $1 - F_X(t)$. The integrand is plotted in Figure 2, which also contains the CDF and SF of a typical lifetime distribution. The area under that integrand, labeled β, is proportional to the bootstrap bias, specifically, the bias is $-\beta/n$. More importantly, if one substitutes the parametric CDF and SF with the respective empirical CDF and SF, $F_n(t)$ and $1 - F_n(t)$, then an empirical estimate for bias can also be calculated. Thus, one must estimate $E(S)$ with $E(S^*)$, producing the empirical estimate of bias as

$$E(S^*) - E(S) \approx -\frac{1}{n}\big(E(S^*) - E(X^*)\big).$$

This estimate of bias has a smaller bias of order $1/n^2$, typically less than one tenth of one percent of a component's expected lifetime for reasonable sample sizes. As we shall see, the empirical bias estimates are within the variability of bias results for various assumptions of parametric distributions.

We can generalize the results of this section for systems consisting of m components in parallel (or series), but the same simplifications do not follow. If we let $S_m^* = \max\{X_1^*, X_2^*, \ldots, X_m^*\}$ be the bootstrap estimate of system lifetime for m components in parallel, then, following a similar approach to the one used in proving Theorem 1, we can find the exact bootstrap bias for a system consisting of m components in

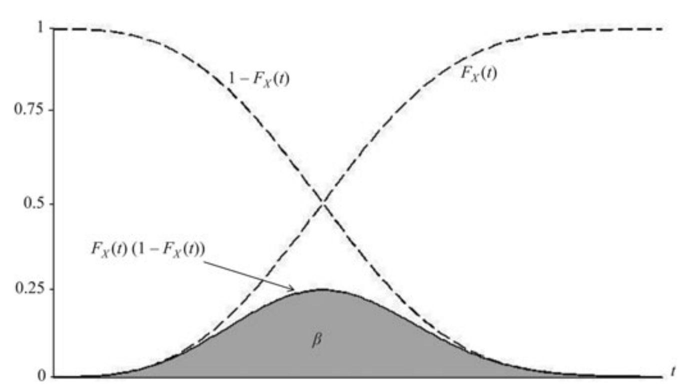

Figure 2. The CDF and SF of a lifetime distribution with its associated bias, which is proportional to the area β.

parallel:

$$\begin{aligned} E(S_m^*) - E(S_m) &= \int_0^\infty t f_{S_m^*}(t)\, dt - \int_0^\infty t f_{S_m}(t)\, dt \\ &= \int_0^\infty \sum_{i=1}^{n} \frac{i^m - (i-1)^m}{n^m} \cdot \frac{n!}{(n-i)!(i-1)!} \\ &\quad t f_X(t) F_X^{i-1}(t)(1-F_X(t))^{n-i}\, dt \\ &\quad - \int_0^\infty t m f_X(t) F_X^{m-1}(t)\, dt. \end{aligned}$$

For example, substituting $m = 3$ and following a similar approach to the one used in Appendix A yields

$$E(S_3^*) - E(S_3) = \frac{(-3n+2)E(S_3) + 3(n-1)E(S_2) + E(X)}{n^2}. \tag{1}$$

We can produce analogous general expressions for bias related to bootstrap estimates for series system lifetimes and letting $m = 3$:

$$E(T_3^*) - E(T_3) = \frac{(-3n+2)E(T_3) + 3(n-1)E(T_2) + E(X)}{n^2}. \tag{2}$$

Empirical estimates for $m = 3$ can be similarly constructed by substituting $E(S_m^*)$ and $E(T_m^*)$ for the appropriate $E(S_m)$ and $E(T_m)$ on the right-hand side of (1) and (2).

Interestingly, assuming different underlying parametric distributions results in fairly consistent bias estimates. Table 2 shows the effect of assuming different distributional families for each component in a two-component parallel system. In each case, the sample size was fixed at $n = 25$ and the distribution's parameters were chosen so that the data had identical population first and second moments μ and σ^2. In other words the different distributions are "similar" at least in "center" and "spread." Note the empirical estimates of bias are within the range of parametric biases, making the concern of the bias' bias less pressing. For the Weibull and gamma distributions, the third and fourth columns of the table represent data with an increasing failure rate (IFR) hazard function, indicative of items that degrade with time, and the last column corresponds to a decreasing failure rate (DFR), indicative of items where "used" is better than "new." In the IFR case the bias is on the order of about 0.1% of the mean failure time. In the DFR case, the biases vary significantly. Even in the worst case (gamma) the bias is about 0.3% of the mean value of the sample. The exponential distribution with means 62, 3, and 20 has associated bootstrap biases −1.24, −0.06, and −0.4, respectively. The exponential distribution lacks the advantage of a second parameter, causing the bias to be of greater magnitude. Note in the Weibull cases where $m = 3$ the bias is greater than for $m = 2$. In general, we have observed that bias increases with m, a conjecture that we have not proven.

5. A PRACTICAL EXAMPLE

In this section we present an example of a fully enumerated bootstrap on the system in Figure 1. The fully enumerated bootstrap (outlined in Efron and Tibshirani 1993) is generally

Table 2. Bootstrapping bias for parametric survival distributions for an m-component parallel system for three population parameter settings and sample size $n = 25$ with μ and σ^2 as shown

Moments and distribution	m	$\mu = 62$ $\sigma^2 = 10$	$\mu = 3$ $\sigma^2 = 0.1$	$\mu = 20$ $\sigma^2 = 2000$
empirical	2	−0.0703	−0.0070	−0.6410
gamma	2	−0.0713	−0.0071	−0.6381
log logistic	2	−0.0696	−0.0069	−0.3705
log normal	2	−0.0713	−0.0071	−0.5249
Weibull	2	−0.0693	−0.0070	−0.5992
Weibull	3	−0.0894	−0.0093	−1.4516

untenable for typical operations involving sums of random variables. Minimum and maximum operations, however, are much more amenable to a fully enumerated bootstrap. Consider, for example, S^* and T^*. The size of their support is limited from above by the total number of distinct observations in the sample **X**. For more complicated coherent systems, the support size of the system bootstrap random variable is similarly constrained because each element in the support of the system lifetime distribution is one of the component lifetimes. The following example illustrates the limited support of RBD systems. The open source Maple-based software *A Probability Programming Language* (APPL) (Drew et al. 2008) allows for such fully enumerated bootstraps of RBDs as well as other systems.

Referring to Figure 1, let the three types of components, A, B, and C, also have previous life test datasets of size $n_A = 45$, $n_B = 60$, and $n_C = 45$, respectively, for each component. The failure times are stored in the Maple lists `alist`, `blist`, and `clist`. Assume that a system designer wants to estimate the lifetime distribution of a system with the series arrangement of parallel subsystems shown in Figure 1. The fully enumerated bootstrap lifetime distribution for this system, $S^* = \min\{\max\{A_1^*, A_2^*\}, \max\{B_1^*, B_2^*\}, \max\{C_1^*, C_2^*\}\}$, can be found with the following APPL commands:

```
A := BootstrapRV(alist);
B := BootstrapRV(blist);
C := BootstrapRV(clist);
Sstar := Minimum(MaximumIID(A, 2),
         Minimum(MaximumIID(B, 2),
         MaximumIID(C, 2)));
```

Note the embedded `Minimum` commands, a necessity since that command only takes on two arguments. The `BootstrapRV` procedure transforms a list into a discrete uniform random variable in APPL's list-of-lists format. For a specific set of data (available in the online supplementary materials), the fully enumerated bootstrapped system was produced and a portion of the 138-element PMF for S^* is as follows:

$$p_{S^*}(t) = \begin{cases} 1/2025, & t = 2.268 \\ 1/675, & t = 3.678 \\ \quad\vdots & \\ 2009/7290000, & t = 10.051 \\ 3599/810000, & t = 11.262 \\ \quad\vdots & \\ 10057/20250000, & t = 173.620 \\ 1157/750000, & t = 179.461. \end{cases}$$

This PMF and the associated CDF are graphed in Figure 3. The graph does not include the vertical lines of a typical plot because they would eclipse some of the points in the plot. Also, each point in the PMF plot has a unique value on the horizontal axis, even though the pattern appears to have points on top of other points. Note how $F_{S^*}(t)$ approximates a smooth curve, that of some underlying, but unknown, continuous $F_S(t)$. Since the PMF of S^* is exactly determined, all the typical characteristics of S^* can be calculated, for example, the mean, variance, skewness, and percentiles. For example $E(S^*) = 78.78$ and the 5th percentile of the system lifetime is 31.95. Also, note that the support size of S^* is limited above by the total number of the individual component failure times, that is, $n_{S^*} = 138 \leq n_A + n_B + n_C = 45 + 60 + 45 = 150$. The algorithm to find $F_{S^*(t)}$ executes quickly, in approximately $O(n_A + n_B + n_C)$ time. In general, for all m, parallel and series systems will execute in $O(n)$ time, as minimum and maximum operations only use the observed failure times. No new values are created that could become support values in S^*. This is unlike a cold backup system, with system lifetime $T = X_1 + X_2$, in which a fully enumerated bootstrap would execute very slowly, on the order of $O(n^2)$, since all $n \times n$ possible support values would need to be enumerated. The authors have done extensive research in these more difficult algebraic structures, creating PMFs with thousands of support values for relatively simple models.

Estimating bias in S^* is possible, but it requires simulation and distributional assumptions. Unlike the case of IID components in series or parallel, there are no easily formed analytic results for non-IID cases, in part due to the innumerable potential configurations of the more complex systems. Using an "estimate then simulate" approach, one can calculate a bias estimate for the system. Making parametric assumptions is helpful in this process, as we will outline next.

Consider the lifetime data available for the individual components A, B, and C. The "estimate" step of the bias estimation process is as follows:

1. Assume a Weibull model for each of the three component types and calculate the associated maximum likelihood estimates $\hat{\eta}_A, \hat{\kappa}_A, \hat{\eta}_B, \hat{\kappa}_B, \hat{\eta}_C$, and $\hat{\kappa}_C$ from each lifetime dataset.
2. Form the estimated CDF for the system

$$\hat{F}_S(t) = 1 - \left[1 - \left(1 - \exp(t^{\hat{\kappa}_A}/\hat{\eta}_A)\right)^2\right] \cdot \left[1-\left(1-\exp(t^{\hat{\kappa}_B}/\hat{\eta}_B)\right)^2\right] \cdot \left[1-\left(1-\exp(t^{\hat{\kappa}_C}/\hat{\eta}_C)\right)^2\right]$$

for $t > 0$.

3. Find the expected system lifetime of the estimated system

$$\hat{E}(S) = \int_0^\infty \left(1 - \hat{F}_S(t)\right) dt.$$

The "simulate" step of the process is as follows:

1. For each i iteration of the simulation, draw n_A simulated component lifetimes from the Weibull($\hat{\eta}_A, \hat{\kappa}_B$) distribution, and likewise n_B and n_C lifetimes from the respective distributions for components B and C.
2. Calculate the empirical distribution for S_i^* from these component lifetimes using the appropriate APPL `Minimum` and `Maximum` commands outlined earlier in this section.
3. Calculate the mean $E(S_i^*)$ for the system.
4. Calculate this iteration's simulated bias, $b_i = E(S_i^*) - \hat{E}(S)$.
5. Repeat this simulation many times, storing each b_i.
6. Find the mean and standard deviation of the bias estimates, $\bar{b}$ and $s_{\bar{b}}$.

This simulation was conducted for 2500 iterations using the values of `alist`, `blist`, and `clist` for the original component lifetimes. In this case $\bar{b} = -0.3553$, $s_{\bar{b}} = 4.513$, and a 95% confidence interval for b was $(-0.5323, -0.1783)$.

To validate that the simulation presents estimates that are reasonable, we have conducted two different evaluations. First, we simulated this "estimate then simulate" process in its own set of simulations. In this effort we established true η and κ parameters for each component. We then simulated 20 sets of life-time data for each component. For each of the 20 sets of data we conducted the "estimate then simulate" process 500 times to produce

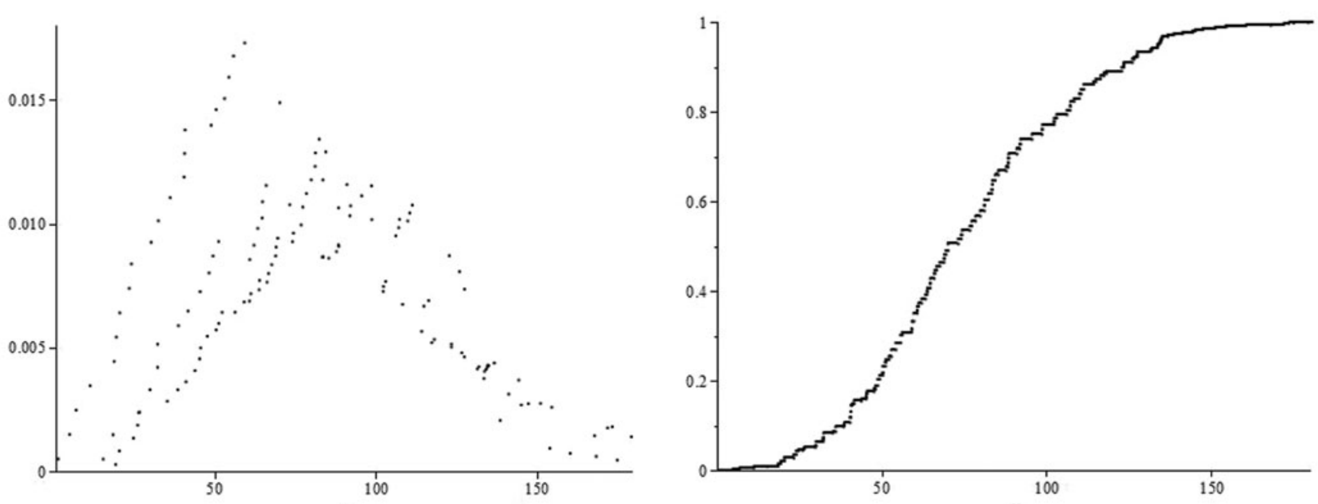

Figure 3. The PMF (left) and CDF (right) of the estimated lifetime of the complex system.

20 values of $\bar{b}$, $\hat{E}(S)$, and 20 confidence intervals. An ANOVA of the 20 × 500 bias values, versus the factor $\hat{E}(S)$ showed that the $\hat{b}$ estimates were statistically similar. We also simulated the effect of reducing the sample sizes for the lifetime data of some of the components. As expected, the bias estimates grew both in expected magnitude and standard deviation when the size of component failure time datasets decreased. This algorithm could be altered to fit specific purposes. For example, one could make the original procedure more robust by averaging over several parametric model assumptions, should that level of confidence be necessary.

The second evaluation of the method involved comparing known bias with simulated bias. We applied this "estimate then simulate" process to IID components in parallel (the subsystems in Figure 1) and compared the simulated bias estimates to the true, analytic biases. In each case the 95% confidence intervals for simulated bias captured the analytic bias. This gives us rationale to opine that the two methods produce stochastically similar bias estimates for these parallel subsystems. Furthermore, the agreement of the analytic and simulated results argues in favor of applying the "estimate then simulate" method to the full system as a way of estimating bias.

In addition to modeling S^* and estimating its bias, a further use of the enumerated bootstrap could be exploring the effect of system design on the system lifetime distribution. Consider again the system in Figure 1. A system designer might want to improve that system by investigating how many components of type A in parallel it would take to increase both the system mean lifetime and the system 5th percentile by, say, 12%. The iterative calculations in APPL and Maple (Maplesoft, 2012), using the code in Appendix C, are used to calculate the answer. It turns out that five components of type A arranged in parallel produce the desired increase of 12% in both the mean and 5th percentile, leaving the designers to then conduct a cost-benefit analysis of the improved system. Further still, one can investigate the asymptotic benefit of even more components of type A in parallel. Let `target` be the desired increase in mean system lifetime. By changing the value of the variable `target` (in Appendix C) in the loop to 1.13, 1.14, 1.15, and 1.16, one finds that the number of type A components in parallel is required to be 6, 7, 20, and ∞, respectively. The asymptotic upper limits of the mean and the 5th percentile are approximately 90.61 and 41.66, respectively, as the number of parallel type A components approaches infinity.

Clearly, without a fully enumerated bootstrap, the above iterative method of exploring system design would become much more computationally expensive. Also, multiple resampling error analyses would have to be conducted at each iteration. In these situations, increasing the bootstrap iteration value B is not as simple as it sounds, as one would have iteration within iteration. One can imagine even more complex optimization algorithms than those above and the benefits of the enumerated bootstrapping become more apparent. Finally, in further research, we have also investigated the cold start backup of components (statistically, distributions of convolutions) using enumerated bootstrapping. Comparing parallel structures of systems to cold start backup structures are now possible, each time getting an exact distribution for the system. In fact all sorts of algebraic operations are possible for other types of systems as well as statistics, many of which we are still investigating. Work on this is also being extended to k-out-of-n type systems as well as cold backup systems.

6. CONCLUSION

Using bootstrapping on a parallel (or series) systems with statistically identical components underestimates (or overestimates) the expected system lifetime. An expression for the bias for all values of n is closed-form. Exact bias results are calculated for two- and three-component systems. Fully enumerated bootstraps are possible, and an application of such bootstrapping is provided.

APPENDIX A

We restate and prove Theorem 1, the bias associated with the bootstrapping for the two-component parallel case. A similar proof construction applies for Theorem 2, the series case, and the more general result for m items in parallel or series. This extra derivation is available in the online supplementary materials section of the journal.

Theorem A.1. For a system consisting of two independent, statistically identical components with finite population means arranged in parallel, the bias resulting from the bootstrap estimation of the mean system lifetime is

$$E(S^*) - E(S) = -\frac{E(S) - E(X)}{n}.$$

Proof. The conditional CDF of X^*, given a random sample $\mathbf{X}$, is the EDF

$$F_{X^*|\mathbf{X}}(t|\mathbf{x}) = \begin{cases} 0 & t < x_{(1)} \\ i/n & x_{(i)} \le t < x_{(i+1)},\ i = 1, 2, \ldots, n-1 \\ 1 & t \ge x_{(n)}. \end{cases}$$

The conditional CDF of $S^* = \max\{X_1^*, X_2^*\}$ is

$$\begin{aligned} F_{S^*|\mathbf{X}}(t|\mathbf{x}) &= P(\max\{X_1^*, X_2^*\} \le t|\mathbf{X} = \mathbf{x}) \\ &= P(X_1^* \le t|\mathbf{X} = \mathbf{x}) \times P(X_2^* \le t|\mathbf{X} = \mathbf{x}) \\ &= F^2_{X^*|\mathbf{X}}(t|\mathbf{x}) \\ &= \begin{cases} 0 & t < x_{(1)} \\ i^2/n^2 & x_{(i)} \le t < x_{(i+1)},\ i = 1, 2, \ldots, n-1 \\ 1 & t \ge x_{(n)}. \end{cases} \end{aligned}$$

The conditional probability mass function (PMF) of S^*on its support is

$$\begin{aligned} p_{S^*|\mathbf{X}}(t|\mathbf{x}) &= F_{S^*|\mathbf{X}}(t|\mathbf{x}) - \lim_{\tau \to t^-} F_{S^*|\mathbf{X}}(\tau|\mathbf{x}) = \frac{i^2 - (i-1)^2}{n^2} \\ &= \frac{2i - 1}{n^2} \qquad t = x_{(i)},\ i = 1, 2, \ldots, n. \end{aligned}$$

Applying the law of total probability and the result concerning the probability distribution of order statistics associated with a continuous population, the PDF of the bootstrap system lifetime is

$$f_{S^*}(t) = \sum_{i=1}^{n} p_{S^*|\mathbf{X}}(t|\mathbf{x}) f_{X_{(i)}}(t) = \sum_{i=1}^{n} \frac{2i-1}{n^2} \cdot \frac{n!}{(n-i)!(i-1)!} f_X(t) F_X^{i-1}(t)\big(1-F_X(t)\big)^{n-i} \quad \text{(A.1)}$$

on its support. Using the binomial theorem to expand $\big(1-F_X(t)\big)^{n-i}$ and then reordering the sums, we can rewrite (A.1) as

$$\begin{aligned} f_{S^*}(t) &= \sum_{i=1}^{n} \frac{2i-1}{n^2} \cdot \frac{n!}{(n-i)!(i-1)!} f_X(t) F_X^{i-1}(t) \\ &\quad \times \sum_{j=0}^{n-i} \binom{n-i}{j} (-1)^{n-i-j} F_X^{n-i-j}(t) \\ &= \sum_{j=0}^{n-1} \frac{f_X(t) n!}{j! n^2} \sum_{i=1}^{n-j} \frac{(2i-1)(-1)^{n-i-j}}{(n-i-j)!(i-1)!} F_X^{i-1}(t) F_X^{n-i-j}(t) \\ &= \frac{f_X(t)}{n} \sum_{j=0}^{n-1} \frac{(n-1)!(n-j-1)! F_X^{n-j-1}(t)}{j!(n-j-1)!} \\ &\quad \times \sum_{i=1}^{n-j} \frac{(2i-1)(-1)^{n-i-j}}{(n-i-j)!(i-1)!} \\ &= \frac{f_X(t)}{n} \sum_{j=0}^{n-1} \binom{n-1}{j} F_X^{n-j-1}(t) \\ &\quad \times \sum_{i=1}^{n-j} \binom{n-j-1}{n-i-j} (2i-1)(-1)^{n-i-j}. \end{aligned} \quad \text{(A.2)}$$

Making the substitution $\alpha = n-i-j$ in the right-most summation,

$$\begin{aligned} \sum_{i=1}^{n-j} \binom{n-j-1}{n-i-j} (2i-1)(-1)^{\alpha} &= \sum_{\alpha=0}^{n-j-1} \binom{n-j-1}{\alpha} \big(2(n-j-\alpha)-1\big)(-1)^{\alpha} \\ &= -2 \sum_{\alpha=0}^{n-j-1} \binom{n-j-1}{\alpha} \alpha (-1)^{\alpha} \\ &\quad + (2n-2j-1) \sum_{\alpha=0}^{n-j-1} \binom{n-j-1}{\alpha} (-1)^{\alpha}. \end{aligned} \quad \text{(A.3)}$$

Using the binomial expansion of $(1+t)^{n-j-1}$ and its derivative,

$$(1+t)^{n-j-1} = \textstyle\sum_{\alpha=0}^{n-j-1} \binom{n-j-1}{\alpha} t^{\alpha}$$
$$(n-j-1)t(1+t)^{n-j-2} = \textstyle\sum_{\alpha=0}^{n-j-1} \binom{n-j-1}{\alpha} \alpha t^{\alpha},$$

we can simplify the summations for certain values of j in (A.3):

$$\sum_{\alpha=0}^{n-j-1} \binom{n-j-1}{\alpha} (-1)^{\alpha} = \begin{cases} 1 & j = n-1 \\ (1-1)^{n-j-1} = 0 & j = 0, 1, \ldots, n-2, \end{cases}$$

and

$$\begin{aligned} \sum_{\alpha=0}^{n-j-1} \binom{n-j-1}{\alpha} \alpha (-1)^{\alpha} &= (-1) \sum_{\alpha=0}^{n-j-1} \binom{n-j-1}{\alpha} \alpha (-1)^{\alpha-1} \\ &= \begin{cases} -1 & j = n-2 \\ -(n-j-1)(1-1)^{n-j-2} = 0 & j = 0, 1, \ldots, n-1, n-3. \end{cases} \end{aligned}$$

Substituting these results into (A.3), we have

$$\sum_{i=1}^{n-j} \binom{n-j-1}{n-i-j} (2i-1)(-1)^{\alpha} = \begin{cases} 1 & j = n-1 \\ 2 & j = n-2 \\ 0 & j = 0, 1, \ldots, n-3. \end{cases}$$

The PDF of S^* from (2) therefore simplifies to

$$f_{S^*}(t) = \frac{f_X(t)\big(2(n-1)F_X(t)+1\big)}{n} = \frac{2(n-1) f_X(t) F_X(t) + f_X(t)}{n}.$$

Using the identities

$$\begin{aligned} E(X) &= \int_0^{\infty} t f_X(t)\, dt, \\ F_S(t) &= F_X^2(t), \\ E(S) &= \int_0^{\infty} t F_S'(t)\, dt = \int_0^{\infty} 2t f_X(t) F_X(t)\, dt, \end{aligned}$$

gives a general result for $E(S^*)$:

$$E(S^*) = \frac{(n-1)\int_0^{\infty} 2t f_X(t) F_X(t)\, dt + \int_0^{\infty} t f_X(t)\, dt}{n} = \frac{(n-1)E(S) + E(X)}{n}.$$

The bias result follows. □

APPENDIX B

The generalized hypergeometric function, ${}_2F_1(\cdot)$ is the following infinite sum:

$${}_2F_1(a, b, c, z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \cdot \frac{z^n}{n!},$$

where $(\cdot)_n$ is the Pochhammer symbol that is shorthand for

$$(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)}.$$

This series converges when $|z| < 1$. Continuity arguments allow for calculating ${}_2F_1(\cdot)$ at $z = -1$. For a full explanation of the function, see Weisstein (2013).

APPENDIX C

The following code, referred to in Section 5, explores the number of type A components to place in parallel to produce the desired increase in performance in the system.

```
n := 2;
target := 1.12;
m := 78.78;
p5 := 31.95;
mux := m;
pct := p5;
while (mux < target * m or pct < target * p5)
do
   n := n + 1;
   System := Minimum(MaximumIID(A, n),
   Minimum(MaximumIID(B, 2),
       MaximumIID(C, 2)));
   mux := Mean(System);
   pct := IDF(System, 0.05);
   end do;
print(n);
```

SUPPLEMENTARY MATERIALS

The first section of the supplement provides the proof of a series system that is analoguous to the parallel system of the primary article. The second section provides the data sets used in the example in Section 5.

[Received May 2012. Revised May 2014.]

References

Arnold, B., Balakrishnan, N., and Nagaraja, H. (2008), *A First Course in Order Statistics*, Philadelphia: SIAM. [176]

Billingsley, P. (1995), *Probability and Measure* (3rd ed.), New York: Wiley. [175]

Doss, H., and Chiang, Y. (1994), "Choosing the Resampling Scheme When Bootstrapping: A Case Study in Reliability," *Journal of the American Statistical Association*, 89, 298–308. [174]

Drew, J., Evans, D., Glen, A., and Leemis, L. (2008), *Computational Probability: Algorithms and Applications in the Mathematical Sciences*, New York: Springer. [178]

Efron, B., and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall. [174,178]

Leemis, L. (2006), "Lower System Reliability Bounds from Binary Failure Data Using Bootstrapping," *Journal of Quality Technology*, 38, 2–13. [174]

Maplesoft (2012), Version 16. Waterloo, Canada. [177,180]

Meeker, W., and Escobar, L. (1998), *Statistical Methods for Reliability Data*, New York: Wiley. [176]

Weisstein, E. W. (2013), "Generalized Hypergeometric Function," *MathWorld–A Wolfram Web Resource*, available at *http://mathworld.wolfram.com/GeneralizedHypergeometricFunction.html*. [177,182]

Wolfram Research, Inc. (2011), Version 8.0, Champaign, IL. [177]