Choosing The Resampling Scheme When Bootstrapping: A Case Study In Reliability

Hani Doss
Florida State University
Department of Statistics
Tallahassee, Florida 32306–3033

and

Yuang-Chin Chiang
Institute of Statistics
National Tsing Hua University
Hsinchu, 300 Taiwan, Republic of China

April 1991

FSU Technical Report No. 844
AFOSR Series D Report No. 1
AFOSR Report No. 91–260
Choosing The Resampling Scheme When Bootstrapping: A Case Study In Reliability

Hani Doss
Florida State University

Yuang-Chin Chiang
National Tsing Hua University

Abstract

Often when dealing with complex data structures there is no unique way to bootstrap. If the data can be viewed as $U_1, \ldots, U_n$ iid from some distribution $P$, one can bootstrap by resampling the $U$'s. Alternately, one can resample in a more model-based way, i.e. by making use of the structure of the model $P$. A typical example of this is linear regression, in which the data is $(Y_i, X_i), i = 1, \ldots, n$. One can resample the pairs $(Y_i, X_i)$ or one can resample the residuals from a fitted model. This phenomenon arises over quite a wide spectrum of problems, and in many cases the different methods of bootstrapping can give substantially different results. It seems hopeless to come up with a general theory that compares the different ways of bootstrapping. In this paper we study in some detail a certain model that arises in reliability theory in which there are two natural ways to bootstrap. This model is described as follows. Available for testing is a sample of $n$ iid systems each having the same structure of $m$ independent components. Each system is continuously observed until it fails. For each component in each system, either a failure time or a censoring time is recorded. A failure time is recorded if the component fails before or at the time of system failure; otherwise a censoring time is recorded. Thus, the system failure acts as a censoring mechanism on the component lifelengths. In this model, bootstrapping can be carried out in two ways. One can resample $n$ systems at random from the original $n$ systems. Alternatively, one can formally compute the Kaplan-Meier estimates $\hat{F}_1, \ldots, \hat{F}_m$ of the component life distributions $F_1, \ldots, F_m$. One then generates artificial lifelengths from these Kaplan-Meier estimates and from those form artificial data. We show that asymptotically, bootstrapping by either method yields correct answers. Intuitively, one expects the model-based method to outperform the other method. The results of an extensive Monte Carlo study show that this is usually true, with substantial gains possible. However, there are also some cases when the model-based method does worse than the “naive” method.

Key words and phrases: Bootstrap, Kaplan-Meier estimator, coherent system, reliability function, martingale central limit theorem.
1 Introduction and Summary

Often when dealing with complex data structures there are several ways to carry out the bootstrap. Examples of this include the following.

(1) **Linear Regression.** The data \((Y_i, X_i), i = 1, \ldots, n\), follow the model 
\[ Y_i = X_i' \beta + \varepsilon_i, \]
where \(X_i\) is a \(p\)-dimensional vector of covariates, \(\beta\) is a \(p\)-dimensional vector of unknown coefficients, and \(\varepsilon_i\) are iid from an unknown distribution \(F\) on \(\mathbb{R}\) with mean 0. Let \(\hat{\beta}\) be an estimate of \(\beta\), whose variability we wish to assess. Bootstrapping can be carried out in two ways.

Method 1. We resample the pairs \((Y_i, X_i), i = 1, \ldots, n\).

Method 2. We resample the (centered) residuals. More specifically, let 
\[ \hat{\varepsilon}_i = Y_i - X_i' \hat{\beta}, \]
\[ \hat{\varepsilon} = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i, \]
and let \(F_n\) be the empirical cdf of \(\hat{\varepsilon}_1 - \hat{\varepsilon}, \ldots, \hat{\varepsilon}_n - \hat{\varepsilon}\). Let \(\hat{\varepsilon}_i^*\) be iid \(\sim F_n\). We construct \(Y_i^* = X_i' \hat{\beta} + \varepsilon_i^*\), and the resampled data is \((Y_i^*, X_i), i = 1, \ldots, n\).

Wu (1986) gives a critical appraisal of these two methods.

(2) **Cox Model.** The model stipulates that the hazard rate for an individual with covariate vector \(x\) is 
\[ \lambda(t|x) = \lambda_0(t) \exp(\beta'x), \]
where \(\lambda_0\) is a completely unspecified hazard function and \(\beta\) is a vector of unknown regression parameters. Associated with individual \(i\) is a covariate vector \(X_i\), a lifefength \(Y_i\), and a censoring time \(C_i\). We do not observe \(Y_i\) directly, but rather we observe \(T_i = Y_i \wedge C_i\) and \(\delta_i = I(Y_i \leq C_i)\). Thus the data is \((T_i, \delta_i, X_i), i = 1, \ldots, n\). The cumulative hazard function for an individual with covariate \(x\) is 
\[ \Lambda(t|x) = \int_0^t \lambda(s|x)ds = \Lambda_0(t) \exp(\beta'x), \]
where \(\Lambda_0(t) = \int_0^t \lambda_0(s)ds\). Thus, the distribution function of the lifetime of an individual with covariate \(x\) is 
\[ F(t|x) = 1 - \prod_{u \leq t} (1 - \Lambda(du))^{\exp(\beta'x)}. \]
If \(\hat{\beta}\) and \(\hat{\Lambda}_0\) are Cox's (1972) and Breslow's (1972, 1974) estimates of \(\beta\) and \(\Lambda_0\), respectively, then \(F(t|x)\) may be estimated by 
\[ \hat{F}(t|x) = 1 - \prod_{u \leq t} (1 - \hat{\Lambda}(du))^{\exp(\beta'x)}. \]
Assume that the \(C_i\)'s are iid \(\sim F_c\), and let \(\hat{F}_c\) be the Kaplan-Meier estimate of \(F_c\) based on the data \((T_i, \delta_i, X_i), i = 1, \ldots, n\). Bootstrapping may be carried out in two ways.

Method 1: Resample the triples \((T_i, \delta_i, X_i), i = 1, \ldots, n\).

Method 2: Generate \(Y_i^* \sim \hat{F}(t|X_i)\) and \(C_i^* \sim \hat{F}_c\), \(i = 1, \ldots, n\), all variables independent. Form \(T_i^* = Y_i^* \wedge C_i^*\) and \(\delta_i^* = I(Y_i^* \leq C_i^*)\). The resampled data is then \((T_i^*, \delta_i^*, X_i), i = 1, \ldots, n\). (Clearly, several other variations on Method 2 are possible, particularly if the \(C_i\)'s are known constants). A reference for this is Hjort (1985), who developed an asymptotic theory for Method 2.

(3) **Parametric Bootstrap.** Let \(X_1, \ldots, X_n\) be iid from the pdf \(f_\theta\) and let \(\hat{\theta}\) be an estimate of \(\theta\), such as the maximum likelihood estimate. For Method 1 of bootstrapping, our resampled data is a random sample of size \(n\) from the empirical distribution function of \(X_1, \ldots, X_n\). For Method 2, our resampled data is a random sample of size \(n\) from \(f_{\hat{\theta}}\).

A common feature in all these problems is that Method 1 is simpler to carry out: A program to implement it is easier to write and faster to run. The gain in simplicity may be substantial. For instance, in Example 3 above, if the \(X_i\)'s are multivariate observations, Method 2 involves generating random vectors from \(f_{\hat{\theta}}\), which may be
difficult to accomplish. Method 1 is also more robust, in the sense that it makes less use of particular aspects of the model. On the other hand if we know that the postulated model holds exactly, then Method 2 is intuitively more attractive and we expect it to be more accurate.

A problem of general interest, therefore, is to obtain information on the loss of efficiency in using Method 1 instead of Method 2: If the loss is not substantial, it may be outweighed by the gain in robustness.

The purpose of this paper is to obtain some concrete results for a certain model arising in reliability theory, in which there are two ways to bootstrap, neither of which is immediately preferable to the other. We next describe this model, give a very brief example of when it arises, and state the results we have obtained for it.

Under study is a coherent structure, denoted by \( \phi \), of \( m \) independent components. Doss, Freitag, and Proschan (1989) (subsequently referred to as DFP) study the problem of estimating \( F \), the distribution of the lifelength of the structure under the following setup. A sample of \( n \) systems, each with the same structure \( \phi \) is available for testing. Each system is continuously observed until it fails. For every component in each system, either a failure time or a censoring time is recorded. A failure time is recorded if the component fails before or at the time of system failure. A censoring time is recorded if the component is still functioning at the time of system failure. This situation arises frequently in reliability theory. It also arises in Quality Control. For example, a standard life testing procedure consists of putting \( m \) randomly sampled items in operation and waiting until the \( r^{th} \) failure. If the \( r \) lifelengths are all recorded then this is Type II censoring (see Lawless (1982), Section 1.4.1) and this corresponds to an \( r\)-out-of-\( m \) system in the model described above. This test is repeated at regular time intervals. Let \( \theta \) be a given parameter of interest and suppose that we wish to detect a change in the value of \( \theta \) as soon as possible. At each time interval \( k \) an estimator \( \hat{\theta}_{[k]} \) of \( \theta \) is made based on the data for the last \( n \) time intervals (\( n \) is adjusted to balance speed of detection of a change and the variance of \( \hat{\theta}_{[k]} \)). When the estimate \( \hat{\theta}_{[k]} \) exceeds certain bounds we declare the value of \( \theta \) to have changed. See e.g. Tomsky (1978).

DFP propose the following procedure. Let \( F_1, F_2, \ldots, F_m \) be the distribution functions of lifelengths of the \( m \) components. Note that we can always relate \( F \) to \( F_1, \ldots, F_m \): There exists a function \( A_\phi : [0,1]^m \rightarrow [0,1] \) such that

\[
F(t) = A_\phi(F_1(t), F_2(t), \ldots, F_m(t)) \quad \text{for} \quad t \geq 0
\]  

(1.1)

(see Chapter 2 of Barlow and Proschan, 1981). For each component, one has data consisting of \( n \) possibly censored survival times, together with an indicator of censoring. Thus, for each \( j \), one can formally compute the Kaplan-Meier estimate \( \hat{F}_j \) of \( F_j \). DFP propose estimating \( F \) by

\[
\hat{F}(t) = A_\phi(\hat{F}_1(t), \hat{F}_2(t), \ldots, \hat{F}_m(t)),
\]

as an alternative to the naive estimate given by the proportion of systems still functioning at time \( t \). (This procedure does not assume that the components are identically distributed, as is the case for the Type II censoring example described above;
however it can be modified if there are "cliques" of identically distributed components.) DFP obtain the asymptotic distribution of $n^{1/2}(\hat{F} - F)$ by showing that the processes $n^{1/2}(\hat{F}_j - F_j)$ converge in distribution to independent Gaussian processes; see Section 2.3 below.

As in Examples 1, 2, and 3 above, in this model there are two ways to bootstrap. Method 1 involves resampling $n$ systems at random from the original $n$ systems. We shall refer to this as the simple method. Method 2 involves resampling by generating independent random lifelengths from the Kaplan-Meier estimates $\hat{F}_j$. More specifically, we generate $X_j^* \sim \hat{F}_j$ independently for $j = 1, \ldots, m$; compute $T^*$, the lifelength of a system whose $j^{th}$ component has lifelength $X_j^*$; then censor $X_j^*$ by $T^*$. The result is artificial data for one system. This is repeated independently $n$ times to obtain an artificial data set of $n$ systems. We call this the obvious method.

Efron (1981) discussed bootstrapping in the random censorship model of survival analysis. This corresponds to a special case of our model, namely the case of a series system of two components. Efron noted that there are two methods of bootstrapping which he called the simple and the obvious methods (our terminology is taken from his) and showed that distributionally, they are the same. In our model, the two methods are not the same: The obvious method can give rise to systems which are distinct from all of the original $n$ systems. The advantage of the obvious method is that it follows the model more closely; in particular it makes use of the assumption of independence of the component lifelengths. The advantage of the simple method is that it is more robust if the independence assumption is violated. It is also faster computationally.

Consider now an arbitrary data set $D$ generated according to some probability mechanism $P$. Suppose that $\eta = \eta(D, P)$ is a function of both the data and the unknown distribution $P$, and suppose that we are interested in estimating $K$, the distribution of $\eta$. The idea of the bootstrap (see Efron and Tibshirani (1986)) is to write $K = K(\hat{P})$, and to estimate $K$ by $K(\hat{P})$, where $\hat{P}$ is an estimate of $P$. To calculate $K(\hat{P})$ we artificially generate data $D^* \sim \hat{P}$ and form $\eta^* = \eta(D^*, \hat{P})$. The distribution of $\eta^*$ is precisely $K(\hat{P})$. Different estimates $\hat{P}$ lead to different methods of bootstrapping. In practice we generate a large number of iid copies $\eta_1^*, \ldots, \eta_B^*$ of $\eta^*$ and use their empirical distribution as our estimate of $K(\hat{P})$. Turning now to asymptotics, suppose that the random quantity $\eta$ has a limiting distribution $K_\infty$. If the distributions of $\eta^*$ converge a.s. to $K_\infty$ (i.e. $K(\hat{P}) \overset{d}{\rightarrow} K_\infty$ a.s.) we shall say that the bootstrap is strongly consistent for $K_\infty$. We shall also use the somewhat less accurate expression "consistent for $\eta".

For the linear regression model (Example 1) Freedman (1981) showed that if $\hat{\beta}$ is the ordinary least squares estimate, the bootstrap is consistent for $\eta = n^{1/2}(\hat{\beta} - \beta)$, whether it is carried out by Method 1 or Method 2. For the standard random censorship model of survival analysis, Akritas (1986), Lo and Singh (1986), and Horváth and Yandell (1987) have, by entirely different techniques, established strong consistency of the bootstrap for $\eta = n^{1/2}(\hat{F} - F)$, $\hat{F}$ being the Kaplan-Meier estimate of $F$.  

3
In this paper we show (Theorems 1 and 2 of Section 2) that for
\[ \eta^{(1)} = n^{1/2}(\hat{F}_1 - F_1, \ldots, \hat{F}_m - F_m) \] and \[ \eta^{(2)} = n^{1/2}(\hat{F} - F) \], the bootstrap is strongly consistent whether it is carried out via the simple or the obvious method. We hope that the techniques we use are also applicable in other problems involving bootstrapping with censored data, such as the Cox model discussed above.

Whereas our asymptotic results prove that either method gives the right answer asymptotically, our Monte Carlo studies indicate that there can be a substantial difference in the mean squared error of the two estimators. For some random variables \( \eta \), Method 2 does better, as one would intuitively expect, and in some cases the gain is substantial. But for other random variables \( \eta \), Method 2 is inferior to the "naive" bootstrap. We have no explanation for this.

This paper is organized as follows. Section 2 describes in detail the setup under study, and gives statements of the main theoretical results of the paper. Section 3 reports the results of our simulation studies. The Appendix gives proofs of the main theoretical results of the paper.

2 Notation, Preliminaries, and Theoretical Results

2.1 Notation

We first introduce the notation needed for a brief review of those results of DFP relevant to the present study. Let

- \( X_{ij} \) = the lifelength of component \( j \) in system \( i \);
- \( T_i \) = the lifelength of system \( i \);
- \( Z_{ij} = \min(X_{ij}, T_i) \);
- \( \delta_{ij} = I(X_{ij} \leq T_i) \);
- \( F_j \) be the distribution function of \( X_{ij} \);
- \( F \) be the distribution function of \( T_i \);
- \( H_j \) be the distribution function of \( Z_{ij} \).

In the definitions above the letter \( i \) indexes systems and the letter \( j \) indexes components. Throughout the paper \( i \) ranges from 1 to \( n \), and \( j \) from 1 to \( m \). The random variables \( X_{ij} \) are not observed. We observe only the \( Z_{ij} \)'s and \( \delta_{ij} \)'s.

It is helpful to keep in mind some concrete examples. Figure 1 shows diagrammatically four simple structures, which we will use in Section 3 for our Monte Carlo studies.

Figures 1c and 1d show parallel systems and these do not involve censoring. Figures 1a and 1b (a 2-out-of-3 system) do. For the system in Figure 1a (the subscript \( i \) indexing systems has been suppressed) \( T = X_1 \wedge (X_2 \vee X_3) \), where \( x \wedge y = \min(x, y) \) and \( x \vee y = \max(x, y) \).

Let \( h : [0, 1]^m \rightarrow [0, 1] \) be the reliability function (this is essentially the function \( h \) appearing in (1.1); see Chapter 2 of Barlow and Proschan 1981, for a definition
and details concerning reliability functions). The survival function of the system lifetime is then

$$\hat{F}(t) = h(\hat{F}_1(t), \ldots, \hat{F}_m(t)) \quad \text{for} \quad t \geq 0$$

(2.1)

where for a distribution function $K$, $\tilde{K}(t)$ denotes $1 - K(t)$. In the example given by Figure 1a, $h(u_1, u_2, u_3) = u_1[1 - (1 - u_2)(1 - u_3)]$. The Kaplan-Meier estimator of $F_j$ is

$$\hat{F}_j(t) = 1 - \prod_{i=1}^{n} \left( 1 - \frac{\delta_{ij}}{n - i + 1} \right)^{\hat{F}_i(t)}$$

(2.2)

where $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(n)}$ and $\delta_{(1)}, \delta_{(2)}, \ldots, \delta_{(n)}$ are the $\delta$’s corresponding to $Z_{(1)}, \ldots, Z_{(n)}$, respectively. The estimator of $F$ proposed by DFP is

$$\hat{F}(t) = 1 - h(\hat{F}_1(t), \ldots, \hat{F}_m(t)).$$

(2.3)

### 2.2 The Underlying Model of Independent Censorship (Random Censorship Model)

The classical results concerning the Kaplan-Meier estimator (weak convergence, asymptotic validity of Greenwood’s formula, etc.) are all proved under the assumption that the random variable of interest and the censoring variable are independent. In our
situation, the component lifelengths are censored by the system lifelength, and the independence condition is clearly violated. It is possible, however, to redefine the censoring variables to bypass this difficulty. Let

$$Y_{ij} = \text{lifelength of system } i \text{ if component } j \text{ is immortal.} \quad (2.4)$$

DFP prove the following intuitively obvious result.

**Lemma 2.1 (Independent Censoring Lemma).**

(i) $X_{ij}$ and $Y_{ij}$ are independent; \quad (2.5)

(ii) $X_{ij} \land T_i = X_{ij}\land Y_{ij}$, and $I(X_{ij} \leq T_i) = I(X_{ij} \leq Y_{ij})$; \quad (2.6)

(iii) $Y_{1j}, \ldots, Y_{nj}$ are iid $\sim G_j$, \quad (2.7)

where

$$G_j(t) = 1 - h(\bar{F}_1(t), \ldots, \bar{F}_{j-1}(t), 1, \bar{F}_{j+1}(t), \ldots, \bar{F}_m(t)). \quad (2.8)$$

Parts (i) and (ii) of the lemma state that the $Y_{ij}$'s act as censoring variables for $X_{ij}$'s. This means that for each $j$, the data $\{(Z_{ij}, \delta_{ij}); i = 1, 2, \ldots, n\}$ for component $j$ follows the random censorship model of survival analysis.

We illustrate on the example given by Figure 1b. Consider component $1$. From (2.4) we see that $Y_1 = X_2 \lor X_3$, and one can check directly that (2.5)–(2.8) hold.

### 2.3 Results From DFP

A weak convergence result for $(\hat{F}_1(t), \ldots, \hat{F}_m(t))$ would imply, via the delta method, a corresponding result for $\hat{F}(t)$, the function $h$ (see (2.1) and (2.3)) being continuously differentiable (see, e.g., Lemma 2.1 of DFP). With this in mind, we define

$$I_j(t) = \frac{\partial h}{\partial u_j}(u_1, \ldots, u_m)\bigg|_{u_k = \hat{F}_k(t), k = 1, \ldots, m}.$$  

Let $D[0, \tau]$ be the Skorohod space on $[0, \tau]$ and $D^m[0, \tau]$ be the product metric space. DFP prove the following.

**Proposition 2.1** Assume that $F_1, \ldots, F_m$ are continuous, and let $\tau$ be such that $F_j(\tau) < 1$ for all $j$. Then as $n \to \infty$,

(A) $n^{1/2}(\hat{F}_1 - F_1, \ldots, \hat{F}_m - F_m) \overset{d}{\to} (W_1, \ldots, W_m)$ in $D^m[0, \tau]$,

where $W_1, W_2, \ldots, W_m$ are independent mean zero Gaussian processes with covariances

$$\text{Cov}(W_j(t_1), W_j(t_2)) = \hat{F}_j(t_1)\hat{F}_j(t_2)\int_0^{t_1} \frac{dF_j(u)}{G_j(u-)F_j(u-)\hat{F}_j(u)} \text{ for } 0 \leq t_1 \leq t_2 \leq \tau.$$  

$$\quad (2.9)$$
(B) \( n^{1/2}(\hat{F} - F) \xrightarrow{d} W \) in \( D[0, \tau] \)

where \( W \) is a mean zero Gaussian process with covariance structure

\[
\text{Cov}(W(t_1), W(t_2)) = \sum_{j=1}^{m} I_j(t_1)I_j(t_2)\hat{F}_j(t_1)\hat{F}_j(t_2)\int_0^{t_1} \frac{dF_j(u)}{G_j(u-)} \frac{dF_j(u)}{G_j(u-)} \text{ for } 0 \leq t_1 \leq t_2 \leq \tau.
\]

We write \( F_j(u-) \) and \( G_j(u-) \) instead of \( F_j(u) \) and \( G_j(u) \) in (2.9) and (2.10) so that these formulas will continue to be valid in Theorem 1, where \( F_j \) and \( G_j \) are not assumed continuous.

We remark that Part B of the proposition follows routinely from Part A.

### 2.4 Two Methods of Bootstrapping

We now give a more explicit description of the two methods of bootstrapping. Let \( P \) denote the true underlying probability model that generates the data

\[
(Z, \delta) = \{(Z_{ij}, \delta_{ij}) ; j = 1, \ldots, m, \ i = 1, \ldots, n\}.
\]

Let \( \eta = \eta((Z, \delta), P) \) be a given function of the data and of the model, and let \( K \) be the distribution of \( \eta \). To estimate \( K \), we need an estimate of \( P \), and as was explained in Section 1, different estimates of \( P \) give rise to different methods of bootstrapping.

**The Simple Method.** We take for \( \hat{P} \) the distribution \( \hat{P}^S \) that gives probability \( 1/n \) to each of the \( n \) systems; so we resample \( n \) systems from the original \( n \) systems. This is formally described as follows. Let

\[
\text{Sys}(i) = \{(Z_{i1}, \delta_{i1}), \ldots, (Z_{im}, \delta_{im})\}
\]

be the data given by the \( i \)th system, for \( i = 1, \ldots, n \).

1. Generate \( k_1, \ldots, k_n \) iid from the uniform distribution over the set \( \{1, \ldots, n\} \).
2. Compute \( \eta^S = \eta((\text{Sys}(k_1), \ldots, \text{Sys}(k_n)), \hat{P}^S) \).
3. Repeat steps 1 and 2 independently \( B \) times, obtaining \( \eta^{S_1}, \ldots, \eta^{S_B} \).

**The Obvious Method.** We note that the assumption of independence of the components implies a correspondence of the form \( P \leftrightarrow (F_1, \ldots, F_m) \). Thus, we may take as \( \hat{P} \) the estimate \( \hat{P} \leftrightarrow (\hat{F}_1, \ldots, \hat{F}_m) \), and the algorithm for the obvious method proceeds as follows.

1. For \( i = 1, \ldots, n, \ j = 1, \ldots, m \), generate \( X_{ij}^* \sim \hat{F}_j \), all \( mn \) variables independent.
2. Form

\[
T_i^{*O} = \text{ lifeflengt } \text{ of system for which component } j \text{ has lifeflengt } X_{ij}^*
\]
and

\[(Z_{ij}^*, \delta_{ij}^*O) = ((X_{ij}^*O \land T_i^*O), I(X_{ij}^*O \leq T_i^*O)).\]

This gives

\[(Z^*, \delta^*O) = \{(Z_{ij}^*, \delta_{ij}^*O), j = 1, \ldots, m, i = 1, \ldots, n\},\]

artificial observed data for \(n\) systems.

3 Compute \(\eta^*O = \eta((Z^*, \delta^*O), \hat{P}O)\).

4 Repeat steps 1, 2, and 3 independently \(B\) times, obtaining \(\eta^{*O1}, \ldots, \eta^{*OB}\).

As was mentioned earlier, a series system of two components corresponds to the standard random censorship model of survival analysis. Efron (1981) considered bootstrapping in this setup and found that

for the random censorship model, the simple and the obvious methods are distributionally identical: They yield

the same bootstrap samples with the same probabilities. \hfill (2.12)

2.5 The Censoring Distributions of \(X_{ij}^*\) Under the Two Bootstrap Methods

The discussion below is helpful in understanding the distinction between the two bootstrap methods. It shows that when bootstrapping, associated with component \(j\) of system \(i\) there are pairs of random variables \((X_{ij}^*S, Y_{ij}^*S)\) and \((X_{ij}^*O, Y_{ij}^*O)\) (corresponding to the simple and obvious methods, respectively). It turns out that \(X_{ij}^*S\)

and \(X_{ij}^*O\) both have the same distribution \(\hat{F}_j\), but the distributions of \(Y_{ij}^*S\) and \(Y_{ij}^*O\) are different.

We discuss the obvious method first. The method here is identical to the non-bootstrapped model, except that \(F_1, \ldots, F_m\) are replaced by \(\hat{F}_1, \ldots, \hat{F}_m\). Therefore, the Independent Censoring Lemma applies. For

\[Y_{ij}^*O = \text{lifelength of a system for which the } k\text{th component}
\]

has lifelength \(X_{ij}^*O\) if \(k \neq j\), and \(\infty\) if \(k = j\)

we have

\[X_{ij}^*O \text{ and } Y_{ij}^*O \text{ are independent,}
\]

\[(Z_{ij}^*, \delta_{ij}^*O) = (X_{ij}^*O \land Y_{ij}^*O, I(X_{ij}^*O \leq Y_{ij}^*O)),\]

and

\[Y_{ij}^*O, \ldots, Y_{nj}^*O \text{ are iid } \sim \hat{G}_j^O,
\]

where

\[\hat{G}_j^O(t) = h((\hat{F}_1(t), \ldots, \hat{F}_{j-1}(t), 1, \hat{F}_{j+1}(t), \ldots, \hat{F}_m(t))). \hfill (2.13)\]
The discussion for the simple method is more subtle. We focus attention on component $j$, and apply the Independent Censoring Lemma (before resampling). We have independent random variables

$$X_{ij} \sim F_j, \ Y_{ij} \sim G_j$$  \hspace{1cm} (2.14)

($G_j$ is given by (2.8)) and the data is

$$(Z_{ij}, \delta_{ij}) = ((X_{ij} \wedge Y_{ij}), I(X_{ij} \leq Y_{ij})), \ i = 1, \ldots, n.$$  \hspace{1cm} (2.15)

Let $\hat{G}_j^S$ be the Kaplan-Meier estimate of $G_j$ based on the data (2.15), i.e.,

$$\hat{G}_j^S(t) = \prod_{i: Z_{ij} \leq t} \left( \frac{n - i}{n - i + 1} \right)^{1 - \delta_{ij}}.$$  \hspace{1cm} (2.16)

Now the simple method corresponds to choosing a sample of size $n$ with replacement from the $n$ pairs $(Z_{ij}, \delta_{ij}), \ i = 1, \ldots, n$. But because the setup (2.14) and (2.15) is the random censorship model, Efron's result (2.12) implies that this simple method is identical to the one in which we generate $X_{ij}^*$ iid $\sim \tilde{F}_j, \ Y_{ij}^*$ iid $\sim \tilde{G}_j^S (X_{ij}^* \wedge Y_{ij}^*)$ independent), and form $(Z_{ij}^{*S}, \delta_{ij}^{*S}) = ((X_{ij}^* \wedge Y_{ij}^*), I(X_{ij}^* \leq Y_{ij}^*))$. Note that $\hat{G}_j^O$ and $\hat{G}_j^S$ (equations (2.13) and (2.16)) are not the same.

We remark that both $\hat{G}_j^O$ and $\hat{G}_j^S$ are strongly uniformly consistent estimators of $G_j$ over $[0, \tau]$. This follows from (2.13),(2.16) and the strong uniform consistency of the Kaplan-Meier estimator (see for example Földes, Rejto, and Winter, 1980).

### 2.6 Theoretical Results of the Paper

We introduce some conditions and conventions (C2 is not needed for Theorem 1).

#### Conditions and Conventions.

**C1** \(\tau\) is any positive number such that $F_j(\tau) < 1$ for $j = 1, \ldots, m$.

**C2** The distribution functions $F_1, \ldots, F_m$ are continuous.

**C3** $(W_1, \ldots, W_m)$ and $W$ are a vector of Gaussian processes on $D^m[0, \tau]$ and a Gaussian process on $D[0, \tau]$, respectively, with distributions as specified in Proposition 2.1.

**C4** $\hat{F}_j^{*S}$ and $\hat{F}_j^{*O}$ denote the Kaplan-Meier estimates for component $j$, computed from the data resampled by the simple and obvious methods, respectively. Also

$$\hat{F}_j^{*S}(t) = 1 - h(\hat{F}_1^{*S}(t), \ldots, \hat{F}_m^{*S}(t))$$

and

$$\hat{F}_j^{*O}(t) = 1 - h(\hat{F}_1^{*O}(t), \ldots, \hat{F}_m^{*O}(t)).$$
Theorem 1 (Obvious Method) Assume C1, C3, and C4. For almost every infinite sequence \(\text{Sys}(1), \text{Sys}(2), \ldots\) (see (2.11)), as \(n \to \infty\),
(A) \(n^{1/2}(\hat{F}_1^0 - \hat{F}_1, \ldots, \hat{F}_m^0 - \hat{F}_m) \overset{d}{\to} (W_1, \ldots, W_m)\) in \(D^m[0, \tau]\).
(B) \(n^{1/2}(\hat{F}^0 - \hat{F}) \overset{d}{\to} W\) in \(D[0, \tau]\).

Theorem 2 (Simple Method) Assume C1–C4. For almost every infinite sequence \(\text{Sys}(1), \text{Sys}(2), \ldots\), as \(n \to \infty\),
(A) \(n^{1/2}(\hat{F}_1^S - \hat{F}_1, \ldots, \hat{F}_m^S - \hat{F}_m) \overset{d}{\to} (W_1, \ldots, W_m)\) in \(D^m[0, \tau]\).
(B) \(n^{1/2}(\hat{F}^S - \hat{F}) \overset{d}{\to} W\) in \(D[0, \tau]\).

REMARKS.
1 The substance of the theorems is contained in Part A, which implies weak convergence for smooth mappings of \(\{(\hat{F}_1^{\star k}, \ldots, \hat{F}_m^{\star k}); t \in [0, \tau]\}(k = S, O)\), with \(\{(\hat{F}_1^{\star k}; t \in [0, \tau]\} \) being one prominent example.
2 Let \(\theta_j\) be the function that maps the probability model \(P\) into the component lifethreshold distribution \(F_j\), i.e., \(\theta_j(P) = F_j\). Then, \(\theta_j(\hat{P}^S) = \hat{F}_j = \theta_j(\hat{P}^O)\). (For \(\hat{P}^O\) this is immediate. For \(\hat{P}^S\) this follows from Efron’s result (2.12)). Thus if \(\eta_j = n^{1/2}(\hat{F}_j - F_j)\), then \(\eta_j^{\star k} = n^{1/2}(\hat{F}_j^{\star k} - F_j)\) for \(k = S, O\) and similarly for \(\eta = n^{1/2}(\hat{F} - F)\).

3 Simulation Studies

Let \((Z^{(n)}, \delta^{(n)})\) (refer to Section 2.4) be data from a sample of \(n\) systems, and consider a random variable \(\eta_n = \eta_n((Z^{(n)}, \delta^{(n)}), P)\) with distribution \(K_\lambda\). We are interested in estimating a parameter \(\theta_n = \theta(K_\lambda)\); \(\theta_n\) could be the variance, for instance. Let \(\hat{K}_n^S\) and \(\hat{K}_n^O\) denote the bootstrap estimates of \(K_\lambda\) by the simple and obvious methods, respectively (i.e., \(\hat{K}_n^S = K_\lambda(\hat{P}^S)\) and \(\hat{K}_n^O = K_\lambda(\hat{P}^O)\)). Also let \(\hat{\theta}_n^S = \theta(\hat{K}_n^S)\) and \(\hat{\theta}_n^O = \theta(\hat{K}_n^O)\); these are the bootstrap estimates of \(\theta_n\) by the simple and obvious methods, respectively.

Theorems 1 and 2, which refer to the random variables \(\eta_n^{(1)} = n^{1/2}(\hat{F}_1 - F_1, \ldots, \hat{F}_m - F_m)\) and \(\eta_n^{(2)} = n^{1/2}(\hat{F} - F)\), can be thought of as asserting an asymptotic equivalence of the two bootstrap methods. For example confidence bands for \(\{F(t); t \in [0, \tau]\}\) (see Bickel and Freedman, 1981, Section 4 and in particular their Corollary 4.2) based on the two methods have asymptotically the same width. In fact, the theorems give only “zeroth order” results. For example, let \(t_0\) and \(s_0\) be fixed, let \(\theta_n = P\{n^{1/2}(\hat{F}(t_0) - F(t_0)) \leq s_0\}\) and let \(\theta_\infty = \lim_{n \to \infty} \theta_n\). Theorems 1 and 2 state that \(\hat{\theta}_n^{\star k} \to \theta_\infty\) a.s., for \(k = S, O\). The theorems say nothing about the rate of convergence or about the distributions of \(\hat{\theta}_n^{\star k}, k = S, O\). Thus they make no statement about the efficiency \(\hat{\theta}_n^S\) vs. \(\hat{\theta}_n^O\). The sequence \(E(\hat{\theta}_n^S - \theta_\infty)^2 / E(\hat{\theta}_n^O - \theta_\infty)^2\) can behave in an arbitrary way without contradicting the theorem (\(a \text{fortiori}\), the same is true for the
sequence \( E(\hat{\theta}_n - \theta_n)^2 / E(\hat{\theta}_n - \theta_n)^2 \). It is therefore of interest to study the empirical behavior of the two bootstrap methods, and this is the purpose of this section.

For our simulation studies we chose the following random variables and parameters.

A \( \eta^A_n / t_p = \hat{F}(t_p) \) and \( \theta^A_n / t_p = \text{Var}(\hat{F}(t_p)) \)

where \( t_p \) is the \( p^{th} \) quantile of \( F \). The values of \( p \) are taken to be .05, .10, .25, .40, .50, .60, .75, .90, and .95.

B \( \eta^B_n = n^{1/2}(\bar{T} - \mu) \) and \( \theta^B_n / t_p = P\{n^{1/2}(\bar{T} - \mu) \leq t_p\} \)

where \( \mu = ET \) and \( t_p \) is the \( p^{th} \) quantile of the distribution of \( n^{1/2}(\bar{T} - \mu) \) (i.e. \( \theta^B_n / t_p = p \)). The values of \( p \) are taken to be .05, .10, .25, .40, .50, .60, .75, .90, and .95.

C \( \eta^C_n = \bar{T} \) and \( \theta^C_n = \text{Var}(\bar{T}) \)

We have also studied the situations which are identical to Cases A and C above, except that "variance" is replaced by "standard deviation", and in addition, we varied the parameters of the exponential distributions and used Weibull distributions in some of the studies. The results were qualitatively very similar, so they are not reported here.

We note that the version of \( \eta^B_n \) under the simple method is \( n^{1/2}(\bar{T} - \bar{T}) \), while under the obvious method it is \( n^{1/2}(\bar{T} - \bar{T}) - E^{O}T_1 \). When resampling by the obvious method we have used the version of the Kaplan-Meier estimate that treats the last observation as uncensored, whether or not it is; this version is a true distribution function. The version given by (2.2) gives positive probability to \(+\infty\) if the last observation is censored, and this can cause \( E^{O}T_1 \) to be infinite.

The parameter in case B is the distribution function evaluated at certain points, while for Cases A and C the parameter is the variance, which is less stable. Theorems 1 and 2 apply to the random variable in Case A. (Actually, the two theorems apply to the distribution of this random variable and technically speaking cannot be extended to apply to the variance, since the variance functional is not weakly continuous.) We have included Cases B and C because they involve the random variable \( \eta = n^{1/2}(\bar{T} - \mu) \), to which Theorems 1 and 2 do not apply. Also, \( \bar{T} \) is an estimator of \( \mu \) that does not use the structure of the model and is therefore valid even if the assumption of independence of the components does not hold.

We have studied a large variety of systems. We report our results only for the four systems shown in Figure 1 because our findings for these systems are representative of the results for our larger studies. We will refer to those systems as System 1, 2, 3, and 4 respectively.

The Monte Carlo studies were carried out as follows. For each system and parameter, we generated \( R \) samples of observations \( \{(X_{ij}^{(r)}), i = 1, \ldots, n, j = 1, \ldots, m\} \), \( r = 1, \ldots, R \), reduced the data to \( (Z_{i}^{(n)}, \delta_{i}^{(n)}) \), \( r = 1, \ldots, R \), and calculated \( \hat{\theta}^{Sr} \) and
\( \hat{\theta}^O, \ r = 1, 2, \ldots, R \) by bootstrapping the sample \( B \) times using the simple and the obvious method, respectively. The mean squared errors were then calculated by

\[
\text{mse}(S) = \frac{1}{R} \sum_{r=1}^{R} (\hat{\theta}^S - \theta)^2 \quad \text{and} \quad \text{mse}(O) = \frac{1}{R} \sum_{r=1}^{R} (\hat{\theta}^O - \theta)^2.
\] (3.1)

The ratio \( \text{mse}(S)/\text{mse}(O) \) indicates the relative performance of the two bootstrap methods.

The distribution of the lifelengths of the components throughout the reported studies was taken to be exponential with parameter 1. The sample size \( n \) was taken to be 10, 30, and 80, and this was done for each system and parameter. Details on how the Monte Carlo experiments were carried out are given later in this section.

The results of our studies are summarized in Tables 1, 2, and 3 below. For each parameter, system, and sample size, the tables give the observed ratio \( \text{mse}(S)/\text{mse}(O) \) and a standard error estimate.

For Study A one makes the following general observations.

1. The ratios are greater than 1 except for a few entries, where the difference between the observed ratio and 1 is well within the sampling error.

2. For systems 1, 3, and 4 the biggest improvement of the obvious method over the simple method tends to be in the middle of the distribution. For system 2 the improvement is in the left tail and decreases monotonically as \( p \) increases.

3. Increasing the redundancy increases the improvement. For parallel systems of \( m \) components, the improvement increases uniformly in \( m \) as \( m \) increases. (We observed this effect by studying values of \( m \) other than 3 and 8; these studies are not reported here.) In the middle of the distribution, (.25 quantile to the .75 quantile) the improvement is very substantial for these highly redundant systems, with a ratio reaching a high of 26.8 for the .6 quantile of the distribution for system 4 when \( n = 80 \).

4. In general, there is no clear trend with increasing sample size that can be detected, except that in the middle of the distribution the ratio increases with increasing sample size.

For Study B, the general trend is that the obvious method shows a quite modest gain in the right and left tails and some loss in the middle of the distribution (the ratio of mean squared errors there is about .7). This trend is common over all sample sizes and all systems, including the two parallel systems. Since the parallel systems do not involve censoring, we conclude that the superiority of the naïve bootstrap over the simple bootstrap is not due to the particular version of the Kaplan-Meier estimator that we used. We found the results of this study surprising.

For Study C, the obvious bootstrap outperforms the simple bootstrap, uniformly over all sample sizes and all systems. The gain is generally modest, however, the ratio ranging from 1.01 to 1.18.

Before drawing general conclusions, we make one cautionary remark about overinterpreting the improvement of the obvious over the simple method in Study A, which
Table 1. Ratio of mse(S) to mse(O) for Var(\(\hat{F}(t_p)\)), where \(t_p\) is the \(p^{th}\) quantile of \(F\). For each \(n, p,\) and system, the first entry in the table is the observed ratio of mean squared errors, and the entry in parentheses is the estimated standard error of the observed ratio.

### n = 10 (number of simulations is 4000; number of bootstraps is 800)

<table>
<thead>
<tr>
<th>(p)</th>
<th>System 1</th>
<th>System 2</th>
<th>System 3</th>
<th>System 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>.05</td>
<td>1.000(.029)</td>
<td>1.322(.080)</td>
<td>1.355(.099)</td>
<td>1.559(.158)</td>
</tr>
<tr>
<td>.10</td>
<td>1.008(.022)</td>
<td>1.303(.044)</td>
<td>1.432(.075)</td>
<td>1.563(.106)</td>
</tr>
<tr>
<td>.25</td>
<td>1.066(.035)</td>
<td>1.229(.029)</td>
<td>1.810(.056)</td>
<td>2.162(.088)</td>
</tr>
<tr>
<td>.40</td>
<td>1.304(.063)</td>
<td>1.109(.019)</td>
<td>3.043(.110)</td>
<td>3.594(.139)</td>
</tr>
<tr>
<td>.50</td>
<td>1.557(.076)</td>
<td>1.073(.014)</td>
<td>4.494(.196)</td>
<td>5.253(.243)</td>
</tr>
<tr>
<td>.60</td>
<td>1.494(.072)</td>
<td>1.049(.014)</td>
<td>3.447(.182)</td>
<td>4.436(.255)</td>
</tr>
<tr>
<td>.75</td>
<td>1.219(.043)</td>
<td>1.035(.018)</td>
<td>1.522(.060)</td>
<td>1.741(.076)</td>
</tr>
<tr>
<td>.90</td>
<td>1.078(.027)</td>
<td>1.033(.015)</td>
<td>1.110(.031)</td>
<td>1.127(.033)</td>
</tr>
<tr>
<td>.95</td>
<td>0.996(.034)</td>
<td>1.038(.011)</td>
<td>1.043(.033)</td>
<td>1.038(.034)</td>
</tr>
</tbody>
</table>

### n = 30 (number of simulations is 4000; number of bootstraps is 3000)

<table>
<thead>
<tr>
<th>(p)</th>
<th>System 1</th>
<th>System 2</th>
<th>System 3</th>
<th>System 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>.05</td>
<td>1.001(.029)</td>
<td>1.149(.052)</td>
<td>1.133(.061)</td>
<td>1.279(.071)</td>
</tr>
<tr>
<td>.10</td>
<td>1.000(.029)</td>
<td>1.180(.034)</td>
<td>1.244(.052)</td>
<td>1.388(.060)</td>
</tr>
<tr>
<td>.25</td>
<td>1.070(.035)</td>
<td>1.158(.027)</td>
<td>1.735(.054)</td>
<td>1.881(.065)</td>
</tr>
<tr>
<td>.40</td>
<td>1.524(.067)</td>
<td>1.078(.014)</td>
<td>3.307(.108)</td>
<td>3.611(.118)</td>
</tr>
<tr>
<td>.50</td>
<td>2.845(.127)</td>
<td>1.036(.009)</td>
<td>8.453(.365)</td>
<td>8.509(.365)</td>
</tr>
<tr>
<td>.60</td>
<td>2.466(.117)</td>
<td>1.021(.007)</td>
<td>10.802(.628)</td>
<td>14.639(.781)</td>
</tr>
<tr>
<td>.75</td>
<td>1.346(.049)</td>
<td>0.996(.009)</td>
<td>2.100(.081)</td>
<td>2.483(.105)</td>
</tr>
<tr>
<td>.90</td>
<td>1.114(.032)</td>
<td>0.987(.012)</td>
<td>1.199(.040)</td>
<td>1.210(.039)</td>
</tr>
<tr>
<td>.95</td>
<td>1.074(.034)</td>
<td>0.992(.011)</td>
<td>1.088(.035)</td>
<td>1.066(.034)</td>
</tr>
</tbody>
</table>

### n = 80 (number of simulations is 800; number of bootstraps is 5000)

<table>
<thead>
<tr>
<th>(p)</th>
<th>System 1</th>
<th>System 2</th>
<th>System 3</th>
<th>System 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>.05</td>
<td>1.005(.074)</td>
<td>1.104(.074)</td>
<td>1.122(.089)</td>
<td>1.186(.110)</td>
</tr>
<tr>
<td>.10</td>
<td>1.014(.070)</td>
<td>1.132(.068)</td>
<td>1.232(.095)</td>
<td>1.287(.100)</td>
</tr>
<tr>
<td>.25</td>
<td>1.064(.071)</td>
<td>1.101(.049)</td>
<td>1.712(.120)</td>
<td>1.912(.139)</td>
</tr>
<tr>
<td>.40</td>
<td>1.645(.136)</td>
<td>1.051(.022)</td>
<td>3.426(.242)</td>
<td>3.335(.246)</td>
</tr>
<tr>
<td>.50</td>
<td>4.223(.323)</td>
<td>1.036(.013)</td>
<td>10.697(.823)</td>
<td>9.306(.801)</td>
</tr>
<tr>
<td>.60</td>
<td>2.843(.252)</td>
<td>1.026(.010)</td>
<td>15.671(1.790)</td>
<td>26.797(2.512)</td>
</tr>
<tr>
<td>.75</td>
<td>1.410(.101)</td>
<td>1.005(.011)</td>
<td>2.057(.168)</td>
<td>3.062(.235)</td>
</tr>
<tr>
<td>.90</td>
<td>1.134(.079)</td>
<td>0.993(.015)</td>
<td>1.160(.080)</td>
<td>1.156(.082)</td>
</tr>
<tr>
<td>.95</td>
<td>1.103(.075)</td>
<td>0.989(.018)</td>
<td>1.099(.071)</td>
<td>1.055(.071)</td>
</tr>
</tbody>
</table>
Table 2. Ratio of \( \text{mse}(S) \) to \( \text{mse}(O) \) for \( P\{n^{1/2}(\bar{T} - \mu) \leq t_p \} \), where \( t_p \) is the \( p^{th} \) quantile of the distribution of \( n^{1/2}(\bar{T} - \mu) \). For each \( n, p, \) and system, the first entry in the table is the observed ratio of mean squared errors, and the entry in parentheses is the estimated standard error of the observed ratio.

<table>
<thead>
<tr>
<th></th>
<th>System 1</th>
<th>System 2</th>
<th>System 3</th>
<th>System 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 10 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p  )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.05</td>
<td>1.102(0.056)</td>
<td>1.230(0.061)</td>
<td>0.964(0.049)</td>
<td>1.017(0.054)</td>
</tr>
<tr>
<td>.10</td>
<td>1.024(0.030)</td>
<td>1.052(0.030)</td>
<td>0.992(0.031)</td>
<td>1.063(0.034)</td>
</tr>
<tr>
<td>.25</td>
<td>0.929(0.026)</td>
<td>0.878(0.024)</td>
<td>1.041(0.029)</td>
<td>1.091(0.031)</td>
</tr>
<tr>
<td>.40</td>
<td>0.903(0.033)</td>
<td>0.816(0.029)</td>
<td>0.951(0.032)</td>
<td>0.986(0.033)</td>
</tr>
<tr>
<td>.50</td>
<td>0.789(0.025)</td>
<td>0.732(0.023)</td>
<td>0.756(0.023)</td>
<td>0.750(0.023)</td>
</tr>
<tr>
<td>.60</td>
<td>0.779(0.027)</td>
<td>0.748(0.025)</td>
<td>0.747(0.025)</td>
<td>0.840(0.028)</td>
</tr>
<tr>
<td>.75</td>
<td>0.924(0.030)</td>
<td>0.872(0.028)</td>
<td>1.023(0.033)</td>
<td>1.062(0.034)</td>
</tr>
<tr>
<td>.90</td>
<td>1.022(0.023)</td>
<td>0.985(0.024)</td>
<td>1.001(0.025)</td>
<td>1.019(0.025)</td>
</tr>
<tr>
<td>.95</td>
<td>1.089(0.039)</td>
<td>1.111(0.041)</td>
<td>0.965(0.038)</td>
<td>1.000(0.037)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>System 1</th>
<th>System 2</th>
<th>System 3</th>
<th>System 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 30 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p  )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.05</td>
<td>1.086(0.046)</td>
<td>1.051(0.041)</td>
<td>0.989(0.042)</td>
<td>0.989(0.046)</td>
</tr>
<tr>
<td>.10</td>
<td>1.022(0.031)</td>
<td>0.990(0.029)</td>
<td>0.992(0.031)</td>
<td>0.986(0.033)</td>
</tr>
<tr>
<td>.25</td>
<td>0.910(0.028)</td>
<td>0.880(0.027)</td>
<td>0.963(0.029)</td>
<td>0.953(0.029)</td>
</tr>
<tr>
<td>.40</td>
<td>0.785(0.025)</td>
<td>0.760(0.025)</td>
<td>0.845(0.027)</td>
<td>0.831(0.027)</td>
</tr>
<tr>
<td>.50</td>
<td>0.672(0.021)</td>
<td>0.628(0.019)</td>
<td>0.694(0.022)</td>
<td>0.671(0.021)</td>
</tr>
<tr>
<td>.60</td>
<td>0.720(0.023)</td>
<td>0.704(0.023)</td>
<td>0.744(0.024)</td>
<td>0.730(0.023)</td>
</tr>
<tr>
<td>.75</td>
<td>0.877(0.028)</td>
<td>0.846(0.028)</td>
<td>0.957(0.030)</td>
<td>0.988(0.032)</td>
</tr>
<tr>
<td>.90</td>
<td>0.975(0.026)</td>
<td>0.949(0.026)</td>
<td>0.990(0.027)</td>
<td>1.028(0.031)</td>
</tr>
<tr>
<td>.95</td>
<td>1.032(0.034)</td>
<td>1.016(0.033)</td>
<td>0.995(0.034)</td>
<td>1.017(0.039)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>System 1</th>
<th>System 2</th>
<th>System 3</th>
<th>System 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 80 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( p  )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.05</td>
<td>1.109(0.088)</td>
<td>1.090(0.079)</td>
<td>0.983(0.079)</td>
<td>1.031(0.084)</td>
</tr>
<tr>
<td>.10</td>
<td>1.050(0.074)</td>
<td>1.035(0.068)</td>
<td>0.965(0.068)</td>
<td>1.026(0.072)</td>
</tr>
<tr>
<td>.25</td>
<td>1.014(0.075)</td>
<td>0.956(0.067)</td>
<td>0.939(0.068)</td>
<td>0.984(0.068)</td>
</tr>
<tr>
<td>.40</td>
<td>0.953(0.069)</td>
<td>0.869(0.061)</td>
<td>0.839(0.058)</td>
<td>0.839(0.061)</td>
</tr>
<tr>
<td>.50</td>
<td>0.680(0.046)</td>
<td>0.701(0.051)</td>
<td>0.673(0.047)</td>
<td>0.633(0.047)</td>
</tr>
<tr>
<td>.60</td>
<td>0.728(0.055)</td>
<td>0.747(0.051)</td>
<td>0.815(0.059)</td>
<td>0.833(0.060)</td>
</tr>
<tr>
<td>.75</td>
<td>0.912(0.068)</td>
<td>0.900(0.066)</td>
<td>0.992(0.073)</td>
<td>0.938(0.063)</td>
</tr>
<tr>
<td>.90</td>
<td>0.973(0.067)</td>
<td>0.946(0.062)</td>
<td>1.031(0.070)</td>
<td>1.049(0.070)</td>
</tr>
<tr>
<td>.95</td>
<td>1.010(0.077)</td>
<td>1.011(0.067)</td>
<td>1.044(0.079)</td>
<td>1.059(0.082)</td>
</tr>
</tbody>
</table>
Table 3. Ratio of $\text{mse}(S)$ to $\text{mse}(O)$ for $\text{Var}(\hat{T})$. For each $n$ and system, the first entry in the table is the observed ratio of mean squared errors, and the entry in parentheses is the estimated standard error of the observed ratio.

<table>
<thead>
<tr>
<th>$n$</th>
<th>No. of simulations</th>
<th>No. of bootstraps</th>
<th>System 1</th>
<th>System 2</th>
<th>System 3</th>
<th>System 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>4000</td>
<td>400</td>
<td>1.116(0.085)</td>
<td>1.182(0.106)</td>
<td>1.013(0.062)</td>
<td>1.055(0.090)</td>
</tr>
<tr>
<td>30</td>
<td>4000</td>
<td>400</td>
<td>1.059(0.060)</td>
<td>1.145(0.061)</td>
<td>1.017(0.058)</td>
<td>1.050(0.055)</td>
</tr>
<tr>
<td>80</td>
<td>4000</td>
<td>400</td>
<td>1.086(0.055)</td>
<td>1.024(0.040)</td>
<td>1.037(0.046)</td>
<td>1.049(0.051)</td>
</tr>
</tbody>
</table>

involves estimation of the variance of $\hat{F}(t)$. The largest ratios of mean squared errors occurred near the middle of the distribution, where these mean squared errors are extremely small. If the main objective is to form confidence intervals, then one should note that the two methods of forming standard confidence intervals (i.e., those of the form estimate $\pm 1.96 \times$ (standard error estimate)) give essentially identical intervals.

Overall, these studies appear to indicate that the obvious method should be preferred to the simple method: It usually results in a gain, and the gain can be very substantial; when it results in a loss, the loss is relatively small. However, the results of Study B give a warning that the model-based bootstrap can be outperformed by the naive bootstrap.

Computational Details

Computations were carried out using Fortran programs compiled with the f77 Unix compiler on the Florida State University Statistics Department network of Sun computers. The simulations were run on several Sparcstation 1's. Pseudo-random uniformly distributed variables were obtained with a Fortran implementation of the "universal random number generator" of Marsaglia and Zaman (1987). Exponential variables were generated by the ziggurat method of Marsaglia and Tsang (1984).

In all the studies we had to obtain the true value of the parameter $\theta$, so that the mean squared errors in (3.1) could be obtained. For Study A, the parameter is $\text{Var}(\hat{F}(t_p))$, and this was obtained by taking the empirical variance of 200,000 independent copies of $\hat{F}(t_p)$. For Study B, the parameter is $P\{n^{1/2}(\hat{T} - \mu) \leq t_p\}$ which is known to be $p$. However, to obtain the bootstrap estimates we must know $t_p$. This was obtained by taking the empirical $p^{th}$ quantile of 500,000 independent copies of $n^{1/2}(\hat{T} - \mu)$. For Study C, the parameter is $\text{Var}(\hat{T})$ which is $(1/n)\text{Var}(T_1)$, and this last variance was obtained by numerical integration (as was the quantity $\mu$ in Study B).

We now discuss the number of bootstraps that were used in obtaining the bootstrap estimates in Study A. In the early stages of this study, our results were ex-
tremely unstable, and we noticed that our estimates varied quite a bit when we increased the number of bootstraps, especially for systems 3 and 4 when \( n \) was 30 or 80. It seemed that an extremely large number of bootstraps was needed to get stable results in the middle of the distribution.

Let \( \hat{\theta}_B \) be a bootstrap estimate of the parameter \( \theta \) based on \( B \) bootstraps, and let \( \hat{\theta}_\infty \) be the “ideal” bootstrap estimate, i.e. the estimate that would be obtained from an infinite number of bootstraps. We then have the following decomposition.

\[
E(\hat{\theta}_B - \theta)^2 = E((\hat{\theta}_B - \hat{\theta}_\infty) + (\hat{\theta}_\infty - \theta))^2
= E(\hat{\theta}_B - \hat{\theta}_\infty)^2 + E(\hat{\theta}_\infty - \theta)^2 + 2E(\hat{\theta}_B - \hat{\theta}_\infty)(\hat{\theta}_\infty - \theta) \tag{3.2}
= E(\hat{\theta}_B - \hat{\theta}_\infty)^2 + E(\hat{\theta}_\infty - \theta)^2 + 2E(\hat{\theta}_\infty - \theta)E\{\hat{\theta}_B - \hat{\theta}_\infty|\text{data}\}.
\]

Thus, if given the data \( \hat{\theta}_B \) is unbiased for \( \hat{\theta}_\infty \) (as will usually be the case), the total mean squared error is the sum of mean squared error due to using only \( B \) bootstraps plus mean squared error due to sampling variability of \( \hat{\theta}_\infty \). The problem that arises is basically that if \( E(\hat{\theta}_\infty - \theta)^2 \) is extremely small, then \( E(\hat{\theta}_B - \hat{\theta}_\infty)^2 \) will make up a substantial proportion of the total error in (3.2). Our objective was to determine the value of \( B \) such that the error due to the finiteness of \( B \) account for, say, only 10% of the total error.

We have studied this problem for the random variable in Study A for the case of parallel systems of \( m \) components. The calculations involved are extremely messy. However, it may be useful to describe briefly here the case \( m = 1 \), since it illustrates the main problem. More importantly, the case \( m = 1 \) is of special interest in its own right since it pertains to bootstrapping in the standard binomial situation.

For \( m = 1 \) the situation is simply the following: \( T_1, \ldots, T_n \) is a sample from the distribution function \( F \), \( F_n \) is the empirical distribution function of \( T_1, \ldots, T_n \), and we wish to estimate \( \text{Var}(F_n(t)) = n^{-1}F(t)(1 - F(t)) \). The bootstrap estimate of \( \text{Var}(F_n(t)) \) is exactly \( n^{-1}F_n(t)(1 - F_n(t)) \), but suppose that we didn’t know this, and instead wanted to calculate this quantity by Monte Carlo. Let \( p = F(t) \) so that the parameter we wish to estimate is \( \theta(p) = n^{-1}p(1 - p) \). At \( p = .5 \), \( \theta(p) \) is “easy to estimate” since \( \theta'(5) = 0 \). Thus, in the decomposition (3.2), as \( p \to .5 \), \( E(\hat{\theta}_\infty - \theta)^2 \) accounts for only a small proportion of the total error, and so larger values of \( B \) are required. This effect becomes more pronounced as \( n \) gets larger.

In more detail, consider the bootstrap estimate of \( \theta \) based on \( B \) bootstrap samples. This is obtained as follows. For \( b = 1, \ldots, B \), let \( X^{*(b)} \) be iid from the binomial distribution with parameters \( n \) and \( \hat{p} \), and let \( \hat{p}^{*(b)} = X^{*(b)}/n \). The estimate \( \hat{\theta}_B \) is

\[
\hat{\theta}_B = \frac{1}{B - 1} \sum_{b=1}^{B} (\hat{p}^{*(b)} - \hat{p}^{*(\cdot)})^2
\]

where \( \hat{p}^{*(\cdot)} = \frac{1}{B} \sum_{b=1}^{B} \hat{p}^{*(b)} \). We then have

\[
E(\hat{\theta}_B - \theta)^2 = E(\hat{\theta}_B - \hat{\theta}_\infty)^2 + E(\hat{\theta}_\infty - \theta)^2. \tag{3.3}
\]
One can calculate explicitly the two expectations on the right side of (3.3) using formulas for the first through fourth moments of the binomial distribution (see, for example Johnson and Kotz, 1969, p. 51). The formulas for these two expectations are extremely messy, but we can give expressions that show the order of magnitude of the terms involved. If $p$ is bounded away from $0$ and $1$, we show below that

$$E(\hat{\theta}_B - \hat{\theta}_{(\infty)})^2 = \frac{1}{Bn^2} \left( 2p(1-p) \right)^2 + \frac{1}{n^2} O\left( \frac{1}{B} \right)$$

(3.4)

and

$$E(\hat{\theta}_{(\infty)} - \theta)^2 = \frac{1}{n^3} p(1-p)(2p-1)^2 + \frac{1}{n^4} \left( 11p^4 - 22p^3 + 13p^2 - 2p \right) + O\left( \frac{1}{n^5} \right)$$

(3.5)

In (3.5) we have included the term of order $n^{-4}$ because the term of order $n^{-3}$ vanishes for $p = .5$.

To obtain (3.4), write

$$E(\hat{\theta}_B - \hat{\theta}_{(\infty)})^2 = EE\left\{ (\hat{\theta}_B - \hat{\theta}_{(\infty)})^2 \mid \text{data} \right\}.$$  

(3.6)

Since given the data $\hat{\theta}_B$ is unbiased for $\hat{\theta}_{(\infty)}$, we have that the conditional expectation in (3.6) is just the conditional variance of $\hat{\theta}_B$ given the data. This is approximately the same as the conditional variance given the data of $B^{-1} \sum_{i=1}^{B} (\hat{\theta}_B - \hat{\theta})^2$, which is equal to

$$\frac{1}{B} \left\{ E(\hat{\theta}_B - \hat{\theta})^4 - (E(\hat{\theta}_B - \hat{\theta})^2)^2 \right\}.$$  

(3.7)

We now use the formulas for the second and fourth central moments of the binomial distributions to obtain (3.4).

To obtain (3.5) we write

$$E(\hat{\theta}_{(\infty)} - \theta)^2 = E \left( \frac{\hat{\theta}(1 - \hat{\theta})}{n} - \frac{p(1-p)}{n} \right)^2,$$  

(3.8)

use the formulas for the first through fourth absolute moments of the binomial distribution, and collect terms.

From (3.4) and (3.5) we see that for $p$ bounded away from $0$ and $1$

$$\frac{E(\hat{\theta}_B - \hat{\theta}_{(\infty)})^2}{E(\hat{\theta}_{(\infty)} - \theta)^2} \approx \frac{\frac{1}{Bn^2} 2(1-p)^2}{\frac{1}{n^2} p(1-p)(2p-1)^2 + \frac{1}{n^4} \left( 11p^4 - 22p^3 + 13p^2 - 2p \right)}.$$  

(3.9)

If $p = .5$ the ratio is $\frac{2}{3} n^2 / B$ and for this to be less than $.1$ we need $B \geq \frac{20}{3} n^2$. These calculations indicate that for $n = 30$ we need $B \geq 6000$ and for $n = 80$ we need $B \geq 42666$. (The exact expressions for the expectations on the right side of (3.3) gave essentially the same numbers; the point of giving expansions (3.4) and (3.5) is to show that at $p = .5$, $B$ must grow at the rate of $n^2$.)

For a parallel system of $m$ components the same kind of calculations can be made. The minimum of the function $E(\hat{\theta}_B - \hat{\theta}_{(\infty)})^2$ occurs at the $.577$ quantile of
the system life distribution function for \( m = 3 \) and at the .595 quantile for \( m = 8 \) (these are precisely the parameter values that gave instability in our Monte Carlo studies). Values of \( B \) of the same order of magnitude are required for the case of a parallel system of \( m \) components if \( m \) is 3 or 8. We do not produce these calculations here.

For \( n = 30 \) we used 3000 bootstraps. For \( n = 80 \) we used 5000 bootstraps, so the results are not reliable for \( p = .6 \). For the parallel system of 8 components and \( p = .6 \), the observed ratio of mean squared errors was only 8.9 when we used 3000 bootstraps; the actual ratio of mean squared errors is likely to be much higher than the 26.8 observed when we used 5000 bootstraps). For \( n = 80 \) we reduced the number of simulations from 4000 to 800 to make the study feasible.

For Studies B and C, much smaller values of \( B \) were sufficient. This was seen empirically and also through some informal calculations.

**Appendix: Proofs of Theoretical Results**

In this section we prove Part A of Theorems 1 and 2. For both Theorems 1 and 2, Part B follows from Part A and a straightforward argument (see the proof of Theorem 2 of DFP); details are omitted. Throughout, \( E^S, E^O, \text{Var}^S, \text{Var}^O \), etc. denote expectations, variances, etc. under the indicated mode of bootstrapping, and the superscripts \( O \) and \( S \) are dropped whenever this will not cause confusion.

### A.1 Proof of Part A of Theorem 1 (Obvious Method)

Let us first restrict our attention to the (relatively simple) task of considering only the \( j \)th component. By the Independent Censoring Lemma the model for component \( j \) is the standard random censorship model. Akritas (1986) showed consistency of the bootstrap for the Kaplan-Meier estimator in this model, i.e. he showed that (in our notation) with probability one, conditional on \( \{(Z_{ij}, \delta_{ij}); \ i = 1, 2, \ldots \} \), as \( n \to \infty \),

\[
n^{1/2}(\hat{F}_j^* - \bar{F}_j) \xrightarrow{d} W_j \tag{A.1}
\]

in \( D[0, \tau] \). This result by itself does not give any information on the joint behavior of \( \hat{F}_1^*, \ldots, \hat{F}_m^* \).

Akritas used the martingale approach to counting processes (a useful reference for this is the survey paper of Andersen and Borgan, 1985). Roughly, the two main steps of his proof are as follows.

1. For almost every infinite sequence \( \{(Z_{ij}, \delta_{ij}); \ i = 1, 2, \ldots \} \), the process

\[
\hat{Z}_j^*(t) = n^{1/2}\left( \frac{\hat{F}_j^*(t) - \bar{F}_j(t)}{\bar{F}_j(t)} \right) \quad 0 \leq t \leq \tau \tag{A.2}
\]

(conditional on \( \{(Z_{ij}, \delta_{ij}); \ i = 1, \ldots, n\} \)) can be represented as a stochastic integral with respect to a martingale, and is therefore also a martingale (for a more precise statement, see (A.34), (A.33) and the remark following (A.36) below).
2 The martingale (A.2) satisfies a.s. the conditions of a central limit theorem for martingales.

The result (A.1) then follows by an easy argument.

Our main focus, then, is the joint asymptotic behavior of \( \hat{F}_1^*, \ldots, \hat{F}_m^* \). Before proceeding with the proof we discuss certain difficulties that arise from bootstrapping.

Generally, proofs of consistency and asymptotic normality based on martingale methods begin by considering certain counting processes of (in our case) the form

\[
N_{ij}(t) = I(X_{ij} \leq t) \quad t \geq 0.
\] (A.3)

These proofs are considerably facilitated if

the processes \( N_{ij} \) have no common jumps with probability one. (A.4)

Condition (A.4) is satisfied if the \( X_{ij} \)'s have continuous distributions, and is heavily used in DFP to obtain the weak convergence of \( n^{1/2}(\hat{F}_1 - F_1, \ldots, \hat{F}_m - F_m) \). For example, (A.4) implies that \( \sum_{i=1}^{n} N_{ij}(t) \) is counting process since its jumps are of size +1 only. It also implies the orthogonality of the martingales formed from \( \sum_{i=1}^{n} N_{ij}(t) \) and \( \sum_{i=1}^{n} N_{ij}(t) \) for \( j_1 \neq j_2 \). See Section 3.3 of Andersen and Eroğan, 1985. In our setup, we will be considering \( X_{ij}^* \sim \hat{F}_j \) and the discreteness of \( \hat{F}_j \) precludes (A.4). Thus, in order to prove Part A of Theorem 1, we shall follow the proof of Theorem 1 of DFP as closely as possible, subject to circumventing (A.4) (however, the proof given here does not depend on the results obtained in DFP).

To obtain the asymptotic independence of \( \hat{F}_1^*, \ldots, \hat{F}_m^* \) (indeed even the fact that they have a joint asymptotic distribution) it will be necessary to introduce filtrations with respect to which the \( m \) processes (A.34) below are simultaneously martingales. It will turn out that in our proof, the role played by filtrations is central (in contrast to Akritas' proof, where filtrations are not specifically mentioned).

We shall need the following two lemmas. In the notation below, \( B_1 \lor B_2 \) denotes the smallest \( \sigma \)-field containing the \( \sigma \)-fields \( B_1 \) and \( B_2 \).

**Lemma A.1** Let \( \{(m(t), \mathcal{F}_t); \ t \geq 0\} \) be a martingale. Suppose that \( \{A_t; \ t \geq 0\} \) is a filtration such that for all \( t, A_t \) is independent of \( \mathcal{F}_t \). Then \( \{(m(t), \mathcal{F}_t \lor A_t; \ t \geq 0\} \) is a martingale.

**Proof.** For \( s < t \), note that \( \mathcal{F}_t \lor \mathcal{F}_s \) is independent of \( A_s \). Since \( m(t) \) is \( \mathcal{F}_t \)-measurable, equation (3) of page 308 of Chung (1974) implies that \( E\{m(t)|\mathcal{F}_s \lor A_s \} = E\{m(t)|\mathcal{F}_s \} \).

The last quantity is \( m(s) \) by the martingale property, and this proves the lemma.

We shall use the notation \( (m_1, m_2)_{B}(\cdot) \) to represent the predictable covariation process of the martingales \( m_1 \) and \( m_2 \) with respect to the filtration \( \{B_u; \ u \geq 0\} \).

Also, \( A \lor B \) will denote the filtration \( \{A_u \lor B_u; \ u \geq 0\} \).

**Lemma A.2** Let \( \{(m_i(t), \mathcal{F}^i_t); \ t \geq 0\}; \ i = 1, 2 \) be two square integrable martingales, such that \( \mathcal{F}^1_t \) and \( \mathcal{F}^2_t \) are independent for all \( t \). Then
(i) \( \{(m_1(t)m_2(t), \mathcal{F}_t^1 \vee \mathcal{F}_t^2); \ t \geq 0\} \) is a martingale.

In particular,

(ii) \( \langle m_1, m_2 \rangle_{\mathcal{F}_t^1 \vee \mathcal{F}_t^2}(t) = 0 \) for all \( t \).

**Proof.** For \( s < t \),

\[
E\left\{ m_1(t)m_2(t) \mid \mathcal{F}_s^1 \vee \mathcal{F}_s^2 \right\} = E\left\{ E\left\{ m_1(t)m_2(t) \mid \mathcal{F}_t^1 \vee (\mathcal{F}_t^1 \vee \mathcal{F}_s^2) \right\} \mid \mathcal{F}_s^1 \vee \mathcal{F}_s^2 \right\} = E\left\{ m_1(t)E\left\{ m_2(t) \mid \mathcal{F}_t^1 \vee \mathcal{F}_s^2 \right\} \mid \mathcal{F}_s^1 \vee \mathcal{F}_s^2 \right\}. \tag{A.5} \]

Now since \( m_2(t) \in \mathcal{F}_t^2 \) and \( \mathcal{F}_t^1 \vee \mathcal{F}_t^2 \) is independent of \( \mathcal{F}_s^1 \), the inner conditional expectation in (A.5) is equal to \( E\{m_2(t)\mid \mathcal{F}_s^2\} \) (again, see equation (3) of page 308 of Chung, 1974) which is equal to \( m_2(s) \) by the martingale property. Thus, (A.5) is equal to

\[
E\{m_1(t)m_2(s)\mid \mathcal{F}_s^1 \vee \mathcal{F}_s^2\} = m_2(s)E\{m_1(t)\mid \mathcal{F}_s^1 \vee \mathcal{F}_s^2\} = m_2(s)m_1(s), \tag{A.6} \]

the last equality in (A.6) being a consequence of Lemma A.1. This proves (i).

To show Part (ii) we note that \( \langle m_1, m_2 \rangle(\cdot) \) is the unique \( \mathcal{F}^1 \vee \mathcal{F}^2 \)-predictable process such that \( \{(m_1(t)m_2(t) - \langle m_1, m_2 \rangle(t), \mathcal{F}_t^1 \vee \mathcal{F}_t^2); \ t \geq 0\} \) is a local martingale (see Theorem 5.2 of Liptser and Shiryaev, 1977).

We now proceed with the proof, our first objective being to show that the processes \( Q_j^* \) given by (A.34) below are orthogonal martingales with respect to an appropriate filtration. This will be accomplished by representing the \( Q_j^* \)'s as stochastic integrals with respect to martingales \( M_j^{*u} \) (see (A.24) below) which we will show are themselves orthogonal.

In the definitions below the sequence \( \{(Z_{ij}, \delta_{ij}); \ j = 1, \ldots, m, \ i = 1, \ldots, n\} \) is considered fixed; thus all distributions are conditional on it. It is convenient to introduce first processes (defined on \([0, \infty)\)) and filtrations corresponding to the case where we imagine that no censoring occurs. The superscript "\( u \)" denotes "uncensored."

\[
N_{ij}^{*u}(t) = I(X_{ij}^* \leq t); \tag{A.7} \]
\[
N_j^{*u}(t) = \sum_{i=1}^n N_{ij}^{*u}(t); \tag{A.8} \]
\[
V_{ij}^{*u}(t) = I(X_{ij}^* \geq t); \tag{A.9} \]
\[
V_j^{*u}(t) = \sum_{i=1}^n V_{ij}^{*u}(t); \tag{A.10} \]
\[
A_{ij}^{*u}(t) = \int_0^t \frac{V_{ij}^{*u}(s)}{\bar{F}_j(s-)} d\bar{F}_j(s); \tag{A.11} \]
\[
A_j^{*u}(t) = \int_0^t \frac{V_j^{*u}(s)}{\bar{F}_j(s-)} d\bar{F}_j(s-); = \sum_{i=1}^n A_{ij}^{*u}(t); \tag{A.12} \]
\[
M_{ij}^{*u}(t) = N_{ij}^{*u}(t) - A_{ij}^{*u}(t); \tag{A.13} \]
\[
M_j^{*u}(t) = N_j^{*u}(t) - A_j^{*u}(t); = \sum_{i=1}^n M_{ij}^{*u}(t); \tag{A.14} \]

20
\[ F^*_t = \text{completion of } \sigma(N^*_i(s); 1 \leq i \leq n, 1 \leq j \leq m, s \leq t). \]  

(A.15)

Now let \( C^*_i(t) \) be the indicator of whether or not component \( j \) in system \( i \) is under observation at time \( t \):

\[ C^*_i(t) = I(Y^*_i \geq t). \]  

(A.16)

The \( C^*_i(t) \)'s are \( F^*_t \)-predictable (this is clear; for a rigorous proof, see Part (ii) of Proposition A.1 of DFP). They enable us to pass conveniently from the processes (A.7)–(A.14) to their analogues for the case where censoring does occur: We define

\[ N^*_ij(t) = \int_0^t C^*_i(s) dN^*_ij(s); \]  

(A.17)

\[ N^*_i(t) = \sum_{i=1}^{n} N^*_ij(t); \]  

(A.18)

\[ V^*_ij(t) = C^*_i(t)V^*_ij(t); \]  

(A.19)

\[ V^*_i(t) = \sum_{i=1}^{n} V^*_ij(t); \]  

(A.20)

\[ A^*_ij(t) = \int_0^t \frac{V^*_ij(s)}{F^*_j(s)} dF^*_j(s); \]  

(A.21)

\[ A^*_i(t) = \int_0^t \frac{V^*_ij(s)}{F^*_j(s)} dF^*_j(s) \quad (= \sum_{i=1}^{n} A^*_ij(t)); \]  

(A.22)

\[ M^*_ij(t) = N^*_ij(t) - A^*_ij(t); \]  

(A.23)

\[ M^*_i(t) = N^*_i(t) - A^*_i(t) \quad (= \sum_{i=1}^{n} M^*_ij(t)); \]  

(A.24)

We shall establish that

\[ \{(M^*_ij(t), F^*_t); t \geq 0\}, \quad j = 1, \ldots, m \quad \text{are martingales and are orthogonal.} \]  

(A.25)

Let

\[ F^{*(i,j)} = \text{completion of } \sigma(N^{*_ij}(s); s \leq t). \]  

(A.26)

It is well known that

\[ \left\{(M^{*_ij}(t), F^{*(i,j)}); t \geq 0 \right\} \text{ is a martingale} \]  

(A.27)

(see e.g., Theorem 18.2 of Liptser and Shiryayev, 1978). Since \( X^*_ij \) and \( X^*_i'j' \) are independent if \( (i, j) \neq (i', j') \), Part (i) of Lemma A.2 implies that

\[ \left\{(M^{*_ij}(t)M^{*_ij'}(t), F^{*(i,j)} \vee F^{*(i',j')}); t \geq 0 \right\} \text{ is a martingale.} \]  

(A.28)

The independence of the \( mn \) variables \( X^*_ki \) implies, via Lemma A.1, that

\[ \left\{(M^{*_ij}(t)M^{*_ij'}(t), F^*_t); t \geq 0 \right\} \text{ is a martingale} \]  

(A.28)
\( (\mathcal{F}_t^\tau \) is defined by (A.15)), and this immediately gives the orthogonality relationship

\[
\langle M_{ij}^{\ast u}, M_{i'j'}^{\ast u} \rangle_{\mathcal{F}_t} = 0 \quad \text{for all } t \geq 0 \quad \text{if } (i, j) \neq (i', j'). \tag{A.30}
\]

To study the \( M_{ij}^{\ast c} \)'s we note that by (A.23), (A.17), (A.21), (A.19), (A.13), and (A.11), we have

\[
M_{ij}^{\ast c}(t) = \int_0^t C_{ij}^\ast(s) dM_{ij}^{\ast u}(s). \tag{A.31}
\]

The \( \mathcal{F}_t^\tau \)-predictability of the \( C_{ij}^\ast \)'s mentioned earlier implies that the Stieltjes integral in (A.31) is a stochastic integral and is a square integrable martingale (see Theorem 18.7 of Liptser and Shiryaev, 1978). Moreover, we have

\[
\langle M_{ij}^{\ast c}, M_{i'j'}^{\ast c} \rangle_{\mathcal{F}_t} = \int_0^t C_{ij}^\ast(s) C_{i'j'}^\ast(s) d\langle M_{ij}^{\ast u}, M_{i'j'}^{\ast u} \rangle_{\mathcal{F}_t}(s) = 0 \tag{A.32}
\]

for all \( t \geq 0 \) if \( (i, j) \neq (i', j') \), the last equality in (A.32) being a consequence of (A.30). The orthogonality relationship (A.32), together with the bilinearity of \( <, > \) now give (A.25).

We now proceed to give a martingale representation to a suitably normalized version of \( (\hat{F}_j^\tau(t), \ldots, \hat{F}_m^\tau(t)) \). Define the stopped process \( \hat{F}_j^\tau \) on \([0, \infty)\) by

\[
\hat{F}_j^\tau(t) = \hat{F}_j(t \wedge Z_{(n)j}^\tau) \tag{A.33}
\]

(recall that \( Z_{(n)j}^\tau = \max\{Z_{1j}^\tau, \ldots, Z_{nj}^\tau\} \)), and let

\[
Q_j^\tau(t) = n^{1/2} \left( \frac{\hat{F}_j^\tau(t) - \hat{F}_j(t)}{\hat{F}_j^\tau(t)} \right) \quad \text{for } t \geq 0. \tag{A.34}
\]

Let \( \tau_n > 0 \) be any number such that \( \hat{F}_j^\tau(\tau_n) < 1 \). From Theorem 3.1 of Aalen and Johansen (1978) or equation (3.2.12) of Gill (1980) we have the representation

\[
Q_j^\tau(t) = n^{1/2} \int_0^t I(V_j^\ast c(s) > 0) \frac{\hat{F}_j^\tau(s)}{V_j^\ast c(s) \hat{F}_j(s)} dM_j^\ast c(s) \quad \text{for } t \leq \tau_n. \tag{A.35}
\]

Because the integrand in (A.35) is \( \mathcal{F}_t^\tau \)-predictable, (A.35) shows that the \( Q_j^\tau \)'s are stochastic integrals and square integrable martingales, and (A.25) gives

\[
\langle Q_j^\tau, Q_j^\tau \rangle_{\mathcal{F}_t}(t) = 0 \quad \text{for } t \leq \tau_n. \tag{A.36}
\]

We remark that \( Z_j^\tau(t) = n^{1/2}( (\hat{F}_j^\tau(t) - \hat{F}_j(t))/\hat{F}_j(t) ) \) need not be a martingale: The expected value of this quantity is 0 for \( t = 0 \) but need not be 0 for \( t > 0 \), since, as is well-known, the Kaplan-Meier estimator can be biased. Thus, equation A.3 of Akritas (1986) is technically incorrect, and in the definition of \( Q_j^\tau \) we do need \( \hat{F}_j^\tau(t) \)
and not \( \hat{F}_j(t) \). However, since a.s. as \( n \to \infty \) \( \hat{H}_j(t) \to H_j(\tau) < 1 \), it is easy to see that
\[
P^{\bullet} \left\{ \sup_{0 \leq t \leq \tau} \left| \frac{\hat{F}_j^* (t) - \hat{F}_j(t)}{\hat{F}_j(t)} - \frac{\hat{F}_j^* (t) - \hat{F}_j(t)}{\hat{F}_j(t)} \right| \neq 0 \right\} \to 0 \text{ a.s. as } n \to \infty. \tag{A.37}
\]

Moreover, since a.s. as \( n \to \infty \) \( \hat{F}_j(t) \to F_j(\tau) < 1 \), the representation (A.35) is valid over \([0, \tau]\) a.s. for large \( n \). Therefore, the arguments in Akritas (1986) go through. The orthogonality relationship (A.36) together with a simple multivariate extension of the martingale central limit theorem used by Akritas can now be used to give the result.

The ideas in this proof can also be used to prove DFP’s Theorem 1 without the assumption of continuity of \( F_1, \ldots, F_m \).

### A.2 Proof of Part A of Theorem 2 (Simple Method)

There does not appear to be any way to use an approach based on counting processes in order to prove Theorem 2. The vector \((N_{ij}^n(t), \ldots, N_{im}^n(t))\) may be completely determined by one of its components when bootstrapping is done via the simple method, and this precludes a structure of orthogonal martingales. Our proof of Theorem 2 is based on a representation for the vector of Kaplan-Meier estimators (Lo and Singh, 1986), more suitable to resampling via the simple method. Let
\[
H_{j1}(t) = P(Z_{ij} \leq t; \delta_{ij} = 1). \tag{A.38}
\]

**Proposition A.1** Under the conditions of Theorem 2,
\[
\begin{pmatrix}
\frac{\hat{F}_1(t) - F_1(t)}{1 - F_1(t)} \\
\vdots \\
\frac{\hat{F}_m(t) - F_m(t)}{1 - F_m(t)}
\end{pmatrix} = \begin{pmatrix}
\frac{1}{n} \sum_{i=1}^{n} \zeta_{i1}(t) \\
\vdots \\
\frac{1}{n} \sum_{i=1}^{n} \zeta_{im}(t)
\end{pmatrix} + \begin{pmatrix}
r_{n1}(t) \\
\vdots \\
r_{nm}(t)
\end{pmatrix} \tag{A.39}
\]

and
\[
\begin{pmatrix}
\frac{\hat{F}_1^S(t) - F_1(t)}{1 - F_1(t)} \\
\vdots \\
\frac{\hat{F}_m^S(t) - F_m(t)}{1 - F_m(t)}
\end{pmatrix} = \begin{pmatrix}
\frac{1}{n} \sum_{i=1}^{n} \zeta_{i1}^S(t) - \frac{1}{n} \sum_{i=1}^{n} \zeta_{i1}(t) \\
\vdots \\
\frac{1}{n} \sum_{i=1}^{n} \zeta_{im}^S(t) - \frac{1}{n} \sum_{i=1}^{n} \zeta_{im}(t)
\end{pmatrix} + \begin{pmatrix}
r_{11}^S(t) \\
\vdots \\
r_{nm}^S(t)
\end{pmatrix}, \tag{A.40}
\]

where
\[
\zeta_{ij}(t) = \frac{1}{H_j(Z_{ij})} I(Z_{ij} \leq t, \delta_{ij} = 1) - \int_0^t \frac{I(Z_{ij} \geq u)}{H_j^2(u)} dH_{j1}(u),
\]
\[
\zeta_{ij}^S(t) = \frac{1}{H_j(Z_{ij}^S)} I(Z_{ij}^S \leq t, \delta_{ij}^S = 1) - \int_0^t \frac{I(Z_{ij}^S \geq u)}{H_j^2(u)} dH_{j1}(u),
\]

23
\[
\sup_{0 \leq t \leq \tau} |r_{n_j}(t)| = O(n^{-3/4} \log^{3/4} n) \text{ a.s.,} \quad \sup_{0 \leq t \leq \tau} |r_{n_j}^*(t)| = O_p(n^{-3/4} \log^{3/4} n) \text{ a.s.,}
\]
for each \( j = 1, \ldots, m, \ i = 1, \ldots, n. \)

**Proof.** For each component, the Independent Censoring Lemma implies that the underlying model is the random censorship model. The proposition follows by applying Theorem 1 of Lo and Singh (1986) to each component. (The discussion in Section 2.5, in particular the paragraph following (2.4), may be helpful here.)

Note that

\[
\zeta_{ij} \text{ and } \zeta_{ij}^* \text{ are uniformly bounded by } c + c^2, \quad \text{ (A.41)}
\]

and that

\[
(\zeta_{i1}^*(t, \ldots, \zeta_{im}^*(t)) \text{ is a random vector taking on the values}
(\zeta_{i1}(t, \ldots, \zeta_{im}(t)), \ i = 1, \ldots, n, \ \text{ with probability } \frac{1}{n}.
\]

Let

\[
K_j(u) = \int_0^u \frac{1}{H_j(v)} dH_j(v).
\]

(A.43)

Note that \( K_j(u) = \int_0^u [\tilde{G}_j(v)]^{-1}[\tilde{F}_j(v)]^{-2} dF_j(v). \)

**Proposition A.2** Assume the conditions of Theorem 2. For each \( s, t \) in \([0, \tau]\) and \( i = 1, \ldots, n, \ j = 1, \ldots, m, \)

\[
E\zeta_{ij}(t) = 0
\]

(A.44)

\[
E\zeta_{ij}(s)\zeta_{ij}(t) = K_j(s \land t),
\]

(A.45)

and

\[
E\zeta_{ij}(s)\zeta_{ij}(t) = 0 \quad \text{ for } \ j_1 \neq j_2.
\]

(A.46)

**Proof.** Let

\[
M_{ij}^u(t) = I(X_{ij} \leq t) - \int_0^t \frac{I(X_{ij} \geq u)}{\tilde{F}_j(u)} dF_j(u)
\]

(A.47)

and

\[
\mathcal{F}_t = \text{ completion of } \sigma(I(X_{ij} \leq s); \ 1 \leq i \leq n, \ 1 \leq j \leq m, \ s \leq t).
\]

We then have the easy facts that

\[
\{(M_{ij}^u(t), \mathcal{F}_t); \ t \geq 0\} \text{ are martingales}
\]

(A.49)
with
\[
(M_{ij}^u, M_{ij'}^u)_{\mathcal{F}}(t) = \begin{cases} 
\int_0^t \frac{I(X_{ij} \geq u)}{F_j(u)} dF_j(u) & \text{if } (i, j) = (i', j') \\
0 & \text{if } (i, j) \neq (i', j')
\end{cases}
\] (A.50)

(for the case $i, j = (i', j')$ this formula is well known; for $(i, j) \neq (i', j')$ the formula is obtained by applying Lemma A.2 and then Lemma A.1). Consider the $\mathcal{F}_t$-predictable indicators of absence of censoring $C_{ij}(t) = I(Y_{ij} \geq t)$, and the stochastic integrals, $M_{ij}^u(t) = \int_0^t C_{ij}(u) dM_{ij}^u(u)$. We then have that
\[
\{(M_{ij}^u(t), \mathcal{F}_t); t \geq 0\} \text{ are martingales} \tag{A.51}
\]

with
\[
(M_{ij}^u, M_{ij'}^u)_{\mathcal{F}}(t) = \begin{cases} 
\int_0^t \frac{I(Z_{ij} \geq u)}{F_j(u)} dF_j(u) & \text{if } (i, j) = (i', j') \\
0 & \text{if } (i, j) \neq (i', j')
\end{cases}
\] (A.52)

The key to the proof of the proposition is to note that $\zeta_{ij}$ is a stochastic integral:
\[
\zeta_{ij}(t) = \int_0^t \frac{1}{H_j(u)} dM_{ij}(u). \tag{A.53}
\]

It then follows that
\[
\{(\zeta_{ij}(t), \mathcal{F}_t); t \geq 0\} \text{ are martingales} \tag{A.54}
\]

with
\[
(\zeta_{ij}, \zeta_{ij'})(t) = \begin{cases} 
\int_0^t \frac{I(Z_{ij} \geq u)}{H_j(u)F_j(u)} dF_j(u) & \text{if } (i, j) = (i', j') \\
0 & \text{if } (i, j) \neq (i', j')
\end{cases}
\] (A.55)

Equation (A.44) is immediate from (A.54). To obtain (A.45) and (A.46), assume that $s \leq t$ and write
\[
E\zeta_{ij_1}(s)\zeta_{ij_2}(t) = E\{E\{\zeta_{ij_1}(s)\zeta_{ij_2}(t)|\mathcal{F}_s}\}
= E\zeta_{ij_1}(s)E\{\zeta_{ij_2}(t)|\mathcal{F}_s\}
= \zeta_{ij_1}(s)\zeta_{ij_2}(s)
= E\{\zeta_{ij_1}, \zeta_{ij_2}\}(s),
\] (A.56)

where the third equality in (A.56) follows from the martingale property and the last equality from the definition of $\{\zeta_{ij_1}, \zeta_{ij_2}\}(\cdot)$. Equations (A.45) and (A.46) now follow from (A.55) and Fubini's theorem.

The following strong uniform consistency results are similar to the Glivenko-Cantelli Theorem. However, the processes involved are of bounded variation instead of being increasing.

**Lemma A.3** There exists $\Omega_0$ with $P(\Omega_0) = 1$ such that for each $\omega$ in $\Omega_0$, as $n \to \infty$,
\[
\sup_{0 \leq t \leq \tau} \left| \frac{1}{n} \sum_{i=1}^n \zeta_{ij}(t) \right| \to 0 \text{ for all } j, \tag{A.57}
\]
\[
\sup_{0 \leq s, t \leq \tau} \left| \frac{1}{n} \sum_{i=1}^{n} \zeta_{ij}(s) \zeta_{ij}(t) - K_j(s \wedge t) \right| \to 0 \quad \text{for all } j, \quad (A.58)
\]
\[
\sup_{0 \leq s, t \leq \tau} \left| \frac{1}{n} \sum_{i=1}^{n} \zeta_{ij_1}(s) \zeta_{ij_2}(t) \right| \quad \text{for all } j_1 \neq j_2. \quad (A.59)
\]

**Proof.** We will prove only (A.58). The proof of (A.59) is similar and that of (A.57) is easier. Fix \( \epsilon > 0 \), and let \( 0 = t_0 \leq t_1 \leq \ldots \leq t_k = \tau \) be a partition of \([0, \tau]\) such that, for \( 0 \leq p, q, k \),
\[
0 \leq K_j(t_p) - K_j(t_{p-1}) < \epsilon, \quad (A.60)
\]
and
\[
0 \leq H_{j_1}(t_p) - H_{j_1}(t_{p-1}) < \epsilon. \quad (A.61)
\]
This is possible since \( K_j \) and \( H_{j_1} \) are uniformly continuous and increasing. By Proposition 3.2 and the strong law of large numbers there exists an \( \Omega_\epsilon \), depending on the partition chosen, with \( P(\Omega_\epsilon) = 1 \) and such that for each \( \omega \in \Omega_\epsilon \), there is an \( N(\omega) \) such that for all \( n \geq N(\omega) \)
\[
\left| \frac{1}{n} \sum \zeta_{ij}(t_p) \zeta_{ij}(t_q) - K_j(t_p \wedge t_q) \right| < \epsilon \quad (A.62)
\]
and
\[
\left| \frac{1}{n} \sum I(t_p < Z_{ij} \leq t_{p-1}, \delta_{ij} = 1) - (H_{j_1}(t_p) - H_{j_1}(t_{p-1})) \right| < \epsilon \quad (A.63)
\]
for all \( p, q \) with \( 0 \leq p, q \leq k \). (The \( \sum \) throughout this proof is taken as a sum as \( i \) ranges from 1 to \( n \).) Now write \( \zeta_{ij}(t) = a_{ij}(t) - b_{ij}(t) \), where
\[
a_{ij}(t) = \frac{1}{H_j(Z_{ij})} I(0 \leq Z_{ij} \leq t, \delta_{ij} = 1), \quad b_{ij}(t) = \int_t^1 I(Z_{ij} \geq u) \frac{1}{H_j(u)} - dH_{j_1}(u).
\]

Note that \( a_{ij} \) and \( b_{ij} \) are increasing processes. For any \( \omega \in \Omega_\epsilon \),
\[
0 \leq a_{ij}(t) \leq \frac{1}{H_j(\tau)} \quad \text{for } 0 \leq t \leq \tau, \quad (A.64)
\]
\[
0 \leq b_{ij}(t) \leq \frac{1}{H_j(\tau)} \quad \text{for } 0 \leq t \leq \tau, \quad (A.65)
\]
and
\[
0 \leq \frac{1}{n} \sum (b_{ij}(t_p) - b_{ij}(t_{p-1})) \leq \frac{\epsilon}{H_j(\tau)} \quad \text{for all } n. \quad (A.66)
\]
Also, by (A.61) and (A.63)
\[
0 \leq \frac{1}{n} \sum (a_{ij}(t_p) - a_{ij}(t_{p-1})) \leq \frac{2\epsilon}{H_j(\tau)} \quad \text{for all } n \geq N(\omega). \quad (A.67)
\]
Thus if \( s \leq t, t_{p-1} \leq t \leq t_p \) and \( t_{q-1} \leq s \leq t_q \) and if \( n \geq N(\omega) \), we have

\[
\frac{1}{n} \sum \zeta_{ij}(t)\zeta_{ij}(s) - K_j(s) \leq \frac{1}{n} \sum a_{ij}(t_p)a_{ij}(t_q) + \frac{1}{n} \sum b_{ij}(t_p)b_{ij}(t_q)
- \frac{1}{n} \sum a_{ij}(t_{p-1})b_{ij}(t_{q-1}) - \frac{1}{n} \sum b_{ij}(t_{p-1})a_{ij}(t_{q-1}) - K(t_{q-1})
= \frac{1}{n} \sum \zeta_{ij}(t_p)\zeta_{ij}(t_q) - K_j(t_q)
+ \frac{1}{n} \sum a_{ij}(t_p)(b_{ij}(t_q) - b_{ij}(t_{q-1})) + \frac{1}{n} \sum b_{ij}(t_{q-1})(a_{ij}(t_p) - a_{ij}(t_{p-1}))
+ \frac{1}{n} \sum b_{ij}(t_p)(a_{ij}(t_q) - a_{ij}(t_{q-1}))
+ K(t_q) - K(t_{q-1})
\leq \epsilon \left( 2 + \frac{6}{H_j^3(\tau)} \right)
\]

by (A.62), (A.64)--(A.67), and (A.60). Similarly

\[
\frac{1}{n} \sum \zeta_{ij}(t)\zeta_{ij}(s) - K_j(s) \geq -\epsilon \left( 2 + \frac{6}{H_j^3(\tau)} \right).
\]

Since the bound is independent of \( t \) and \( s \), and \( \epsilon \) is arbitrary, (A.58) holds by taking \( \epsilon = \frac{1}{k} \) and \( \Omega_0 = \cap_k \Omega_k \).

**Proof of Theorem 2.** Since the error terms \( r_{n_1}^{*S}, \ldots, r_{n_m}^{*S} \) in Proposition 3.1 converge to zero in (simple) bootstrap probability for all \( \omega \in \Omega_1 \) for some \( \Omega_1 \) with \( P(\Omega_1) = 1 \), we need only show that

\[
n^{1/2} \begin{pmatrix} \tilde{\zeta}_{n_1}^{*S} - \tilde{\zeta}_{n_1} \\ \vdots \\ \tilde{\zeta}_{n_m}^{*S} - \tilde{\zeta}_{n_m} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} W'_1 \\ \vdots \\ W'_m \end{pmatrix}
\]

weakly in \( D^m[0, \tau] \), where

\[
\tilde{\zeta}_{n_j}^{*S}(t) = \frac{1}{n} \sum_{i=1}^n \zeta_{ij}^{*S}(t), \quad \tilde{\zeta}_{n_j}(t) = \frac{1}{n} \sum_{i=1}^n \zeta_{ij}(t),
\]

and \( W'_1, \ldots, W'_m \) are independent Gaussian processes with mean zero and covariance structure

\[
EW'_j(t)W'_j(s) = K_j(t \wedge s).
\]

It is not difficult to see that weak convergence in the product metric space \( D^m[0, \tau] \) follows from tightness and convergence of finite dimensional distributions:

\[
n^{1/2} \begin{pmatrix} (\tilde{\zeta}_{n_1}^{*S}(t_1) - \tilde{\zeta}_{n_1}(t_1)), \ldots, (\tilde{\zeta}_{n_1}^{*S}(t_k) - \tilde{\zeta}_{n_1}(t_k)) \\ \vdots \\ (\tilde{\zeta}_{n_m}^{*S}(t_1) - \tilde{\zeta}_{n_m}(t_1)), \ldots, (\tilde{\zeta}_{n_m}^{*S}(t_k) - \tilde{\zeta}_{n_m}(t_k)) \end{pmatrix}'
\xrightarrow{d} (W'_1(t_1), \ldots, W'_1(t_k), \ldots, W'_m(t_1), \ldots, W'_m(t_k))'
\]

(A.68)
for all \( 0 \leq t_1 \leq t_2 \leq \ldots \leq t_k \leq \tau \). (The superscript ‘ above means transpose of a vector.) The tightness of the vector processes follows immediately from that of the component processes and the definition of tightness. Statement (A.68) is, by the Cramér-Wold device, equivalent to

\[
 n^{1/2} \sum_{j=1}^m \sum_{l=1}^k a_{jl} (\bar{\zeta}_{nj}^S(t_l) - \bar{\zeta}_{nj}(t_l)) \overset{d}{\to} \sum_{j=1}^m \sum_{l=1}^k a_{jl} W_j'(t_l) \tag{A.69}
\]

for all \( a_{jl} \in R, \ j = 1, \ldots, m, \ l = 1, \ldots, k \).

Write the left side of (A.69) as \( n^{-1/2} \sum_{i=1}^n A_{ni} \), where

\[
 A_{ni} = \sum_{j=1}^m \sum_{l=1}^k a_{jl} (\zeta_{ij}^S(t_l) - \zeta_{nj}(t_l)).
\]

We now have a standard triangular array setup to which we intend to apply the Liapounov version of the central limit theorem. We proceed to check the moments.

\[
 E^* A_{ni} = \sum_{j=1}^m \sum_{l=1}^k a_{jl} E^* (\zeta_{ij}^S(t_l) - \zeta_{nj}(t_l)) = 0,
\]

by (A.42). To calculate the variance, we write

\[
 \text{Var}^* A_{ni} = E^* \left[ \sum_{j=1}^m \sum_{l=1}^k a_{jl} (\zeta_{ij}^S(t_l) - \zeta_{nj}(t_l)) \right]^2
\]

\[
 = \sum_{j=1}^m \sum_{l_1=1}^k \sum_{l_2=1}^k a_{jl_1} a_{jl_2} \left[ E^* (\zeta_{ij_1}^S(t_{l_1}) - \zeta_{nj_1}(t_{l_1})) (\zeta_{ij_2}^S(t_{l_2}) - \zeta_{nj_2}(t_{l_2})) \right]
\]

\[
 + \sum_{j_1 \neq j_2} \sum_{l_1=1}^k \sum_{l_2=1}^k a_{jl_1} a_{jl_2} \left[ E^* (\zeta_{ij_1}^S(t_{l_1}) \zeta_{ij_2}(t_{l_1}) - \zeta_{nj_1}(t_{l_1}) \zeta_{nj_2}(t_{l_1})) \right]
\]

\[
 = \sum_{j=1}^m \sum_{l_1=1}^k \sum_{l_2=1}^k a_{jl_1} a_{jl_2} \left( \frac{1}{n} \sum_i \zeta_{ij}(t_{l_1}) \zeta_{ij}(t_{l_2}) - \zeta_{nj}(t_{l_1}) \zeta_{nj}(t_{l_2}) \right)
\]

\[
 + \sum_{j_1 \neq j_2} \sum_{l_1=1}^k \sum_{l_2=1}^k a_{jl_1} a_{jl_2} \left( \frac{1}{n} \sum_i \zeta_{ij_1}(t_{l_1}) \zeta_{ij_2}(t_{l_2}) - \zeta_{nj_1}(t_{l_1}) \zeta_{nj_2}(t_{l_2}) \right),
\]

where the last equality (the heart of the calculation) follows from (A.42). By Lemma A.3, for each \( \omega \in \Omega_o \cap \Omega_1 \), this converges to the constant \( A \) given by

\[
 A = \sum_{j=1}^m \sum_{l_1=1}^k a_{jl_1} \sum_{l_2=1}^k K_j(t_{l_2} \wedge t_{l_1}).
\]

Since by (A.41) \( A_{ni} \) is bounded by a constant \( C \), we have \( \sum_{i=1}^n E^* \left| A_{ni} \right|^3 \overset{n^{3/2}}{\to} 0 \).

The Liapounov Theorem implies that

\[
 n^{-1/2} \sum_{i=1}^n A_{ni} \overset{d}{\to} N(0, A).
\]

But \( N(0, A) \) is the distribution of \( \sum_{j=1}^m \sum_{l=1}^k a_{jl} W_j'(t_l) \). This proves (A.69), (A.68) and the theorem.
References


Choosing The Resampling Scheme When Bootstrapping: A Case Study In Reliability

Hani Doss and Yuang-Chin Chiang

Type of Report: Technical

Time Covered: From April 1991

Subject Terms: Bootstrap, Kaplan-Meier estimator, coherent system, reliability function, martingale central limit theorem.

Abstract: See Back

Distribution/Availability of Abstract: Unclassified/Unlimited

Name of Responsible Individual: Frank Proschan

Telephone Number: (904) 644-3218

Office Symbol: AFOSR/NM