Journal of the Korean
Statistical Society
Vol. 24, No. 1, 1995

Conditional Bootstrap Methods for Censored Survival Data [†]

Ji-Hyun Kim [1]

ABSTRACT

We first consider the random censorship model of survival analysis. Efron (1981) introduced two equivalent bootstrap methods for censored data. We propose a new bootstrap scheme, called Method 3, that acts conditionally on the censoring pattern when making inferences about aspects of the unknown life-time distribution F. This article contains (a) a motivation for this refined bootstrap scheme; (b) a proof that the bootstrapped Kaplan-Meier estimator of F formed by Method 3 has the same limiting distribution as the one by Efron's approach; (c) description of and report on simulation studies assessing the small-sample performance of the Method 3; (d) an illustration on some Danish data. We also consider the model in which the survival times are censored by death times due to other causes and also by known fixed constants, and propose an appropriate bootstrap method for that model. This bootstrap method is a readily modified version of the Method 3.

KEYWORDS: Bootstrap, Kaplan-Meier estimator, Ancillary statistics, Martingale central limit theorem, Stochastic integral.

[†]This research was supported by Soong Sil University Research Fund(1995)
[1] Department of Statistics, Soong-Sil University, Dongjak, Seoul, 156-743, Korea

1. INTRODUCTION

The goal of this article is to study bootstrapping in simple models of survival analysis which involve censoring. We begin by studying the random censorship model, which we now review. Let the random variables $X_1, \ldots, X_n$ be nonnegative independent and identically distributed (iid) with distribution function F and let $Y_1, \ldots, Y_n$ be iid $\sim G$. We assume that the X's and the Y's are independent. The X's represent survival times, the Y's represent censoring times. We observe $T_i = \min(X_i, Y_i)$, $\delta_i = I(X_i \le Y_i)$, $i = 1, \ldots, n$. Clearly $T_1, \ldots, T_n$ are iid $\sim H$, where $1 - H = (1-F)(1-G)$.

Efron (1981) introduced the following two bootstrap methods for censored data and showed that they are distributionally the same.

Method 1. Draw an iid sample of pairs $(T_1^*, \delta_1^*), \ldots, (T_n^*, \delta_n^*)$ from the n pairs $(T_1, \delta_1), \ldots, (T_n, \delta_n)$, in which each pair (T_i^*, δ_i^*) takes the values (T_j, δ_j) with probability $1/n$, $j = 1, \ldots, n$.

Method 2. Generate $X_i^* \sim F_n$, and $Y_i^* \sim G_n$, where F_n and G_n are the Kaplan-Meier estimators (KME, Kaplan and Meier 1958) of F and G respectively. Then form (T_i^*, δ_i^*), where $T_i^* = \min(X_i^*, Y_i^*)$, $\delta_i^* = I(X_i^* \le Y_i^*)$, $i = 1, \ldots, n$.

Let us now consider bootstrapping more closely. Suppose that we wish to estimate the variance of $F_n(t)$. If we knew the Y's then we would condition on them by the ancillary principle, since the distribution of the Y's does not depend on F. That is, we would want to estimate $\mathrm{Var}\{F_n(t) | Y_1, \ldots, Y_n\}$. Unfortunately, in the random censorship model we do not see all the Y's. If $\delta_i = 0$ we see the exact value of Y_i, but if $\delta_i = 1$ we know only that $Y_i > T_i$. Let us denote this information on the Y's by $\mathcal{C}$. Then, what we want to estimate is $\mathrm{Var}\{F_n(t) | \mathcal{C}\}$. Efron's scheme, which is appropriate for estimating the unconditional variance, can be adapted to provide an estimate of $\mathrm{Var}\{F_n(t) | \mathcal{C}\}$, as follows.

Method 3. Draw $X_i^* \sim F_n$ and generate Y_i^* as follows:

- If $\delta_i = 0$, then we know that $Y_i = T_i$, so we take $Y_i^* = T_i$;
- If $\delta_i = 1$, then we know that $Y_i > T_i$, so we generate Y_i^* from the distribution

$$\frac{G_n(t) - G_n(T_i)}{1 - G_n(T_i)} \quad \text{for } t > T_i. \tag{1.1}$$

(Note that conditionally on the observations $(T_1, \delta_1), \ldots, (T_n, \delta_n)$, $G_n(t)$ is a (sub)distribution function and (1.1) is a conditional (sub)distribution function of $G_n(t)$. It also should be noted that when $\delta_i = 1$, $G_n(T_i) < 1$ so that the denominator in (1.1) is not 0.)

Now form $(T_i^*, \delta_i^*), i = 1, \ldots, n$.

In Methods 1 and 2, $T_1^*, \ldots, T_n^*$ are iid H_n with $1 - H_n(t) = (1 - F_n(t))(1 - G_n(t)) = (n - k_t)/n$ where k_t is the number of T_i's that are less than or equal to t, and $\delta_i^* = \delta_j$ if $T_i^* = T_j$ for $i = 1, \ldots, n$. In Method 3, $T_1^*, \ldots, T_n^*$ are independent but not identically distributed since the Y_i^*'s have different distributions for different i's. Thus, Method 1 (2) and Method 3 are different.

In some situations the assumption of the random censorship model where the censoring times $Y_1, \ldots, Y_n$ are iid may not be appropriate. For example we may plan a medical study which is scheduled to end ten years from some starting date. Suppose that there are several possible causes of death and we are particularly interested in one specific cause. And suppose that all patients enter the study along the way over a ten year period. When the i^{th} patient enters the study, the patient's censoring time due to the termination of study will be known and fixed as, say w_i. Let C_i be the censoring time due to death from causes other than the one of interest. Then the overall censoring time would be $Y_i = \min(w_i, C_i)$, and $Y_1, \ldots, Y_n$ are not identically distributed even if we assume $C_1, \ldots, C_n$ are iid $\sim K$. For this setup Bootstrap Method 1 (or 2) is not logically correct to apply since the censoring times do not have a common distribution. But Method 3 can be easily extended to meet this situation. Here we observe $(T_i, w_i, \delta_i), i = 1, \ldots, n$ where w_i is a fixed censoring time for the i^{th} patient, $T_i = \min(X_i, C_i, w_i)$ and $\delta_i = 1$ if $T_i = X_i$ (uncensored), $\delta_i = .5$ if $T_i = C_i$ (censored due to death by other causes), $\delta_i = 0$ otherwise (censored due to the end of study). Let K_n be the KME of K, obtained when we consider C_i's as survival times and consider X_i's and w_i's as censoring times, i.e. K_n is the KME based on the data (T_i, η_i), where $\eta_i = 1$ if $\delta_i = .5$ and $\eta_i = 0$ if $\delta_i = 0$ or 1.

Method 3′ For $i = 1, \ldots, n$ draw $X_i^* \sim F_n$ and generate Y_i^* as follows:

- If $\delta_i \neq 1$, then we know that $Y_i = T_i$, so we take $Y_i^* = T_i$.

- If $\delta_i = 1$, then we know that $Y_i > T_i$. Since w_i is fixed, we only need to generate C_i^* to get Y_i^*, i.e. generate C_i^* from the distribution

$$\frac{K_n(t) - K_n(T_i)}{1 - K_n(T_i)} \quad \text{for } t > T_i, \tag{1.2}$$

We then take $Y_i^* = \min(C_i^*, w_i)$.

Now form (T_i^*, w_i, δ_i^*) from X_i^*, Y_i^* and w_i. More specifically, $\eta_i^* = \min(X_i^*, Y_i^*)$, $\delta_i^* = 1$ if $X_i^* \leq Y_i^*$, $\delta_i^* = .5$ if $Y_i^* < X_i^*$ and $Y_i^* = C_i^*$, and $\delta_i^* = 0$ if $Y_i^* < X_i^*$ and $Y_i^* = w_i$. Note that the fixed censoring times $w_1, \ldots, w_n$ are not resampled.

For the extreme case in which all censoring times are due to the termination of the study (fixed censorship model), Efron (1981) suggested the following approach: choose $X_1^*, \ldots, X_n^* \overset{iid}{\sim} F_n$, and define $T_i^* = \min(X_i^*, w_i)$ with δ_i^* equaling 1 or 0 as T_i^* equals X_i^* or w_i. Our Method 3′ becomes equivalent to this approach when there is no censoring time due to other causes.

Let us state the general idea of the bootstrap (see Efron and Tibshirani, 1986). We have a random quantity of interest $R = \eta(D, P)$, which is a function of both the data D and the unknown probability mechanism P that generates D. We wish to estimate some aspect of the distribution of R. We assume that we have some way of estimating the probability model P from the data D, producing $\hat{P}$. Once we have $\hat{P}$, we can generate D^* from $\hat{P}$ by Monte Carlo methods, so that we observe $R^* = \eta(D^*, \hat{P})$. The idea of the bootstrap is to estimate some aspect of the distribution of R by that of R^*. Different estimates of $\hat{P}$ lead to different methods of bootstrapping. For example, in the random censorship model, we may take $\hat{P} = (\hat{F}, \hat{G})$, where $\hat{F}$ is the estimate of the life-length distribution and $\hat{G}$ is the estimate of the censoring time distribution. For Method 3 and 3′ we have different ways of estimating the model P. Considering now the asymptotics, suppose that the random quantity R has a limiting distribution L_∞. If as $n \to \infty$ the distribution of R^* converges a.s. to L_∞ we shall say that the bootstrap *is consistent for* R. For the standard random censorship model Akritas (1986) has established the consistency of the bootstrap by Method 2 (equivalently Method 1) for $R = n^{1/2}(F_n - F)$, using the theory of martingales for point processes.

In this article we show (Theorems 1 and 2 of Section 3) that for $R = n^{1/2}(F_n - F)$, the bootstrap is consistent whether it is carried out via Method

3 or Method 3′. Whereas these asymptotic results prove that all bootstrap methods are consistent for $R = n^{1/2}(F_n - F)$, they say nothing at all about the rate of convergence or the finite-sample performance of the various bootstrap methods. In Section 2 simulation studies are reported that assess the small-sample performance of the various bootstrap methods. The variance of the KME, confidence bands for the survival curve, confidence intervals for the survival curve at a fixed time and confidence intervals for quantiles are examined. Finally, we apply the various bootstrap methods to real data in Section 4.

2. SIMULATION STUDIES

Let $D^{(n)}$ be the data. For example, $D^{(n)} = ((T_1, \delta_1), \ldots, (T_n, \delta_n))$ for the standard random censorship model. Consider a random quantity $R_n = \eta(D^{(n)}, P)$ with distribution L_n. We wish to estimate a parameter $\theta_n = \theta(L_n)$, some aspect of the distribution of R_n. Let $\widehat{L}_n^i$ denote the bootstrap estimates of L_n by method i for $i = 1, 2, 3, 3'$. Also let $\hat{\theta}_n^i = \theta(\widehat{L}_n^i)$, the bootstrap estimate of θ_n by method i.

For the first simulation study, designated by A, we chose the following random quantity and parameter.

A $R_n^A(t) = F_n(t)$ and $\theta_n^A(t) = \mathrm{Var}(F_n(t))$.

For other simulation studies, designated by B, C and D, we obtained confidence intervals and confidence bands as follows.

B Let $R_n^B = \{n^{1/2}(F_n(t) - F(t)); 0 \leq t \leq \tau\}$ with τ defined in Theorem 1 in Section 3. We obtained a confidence band for F on $[0, \tau]$ by obtaining a real number U_α such that $P\{\sup_{0\leq t\leq \tau} n^{1/2}|F_n(t) - F(t)| \leq U_\alpha\} = 1 - \alpha$.

C Let $R_n^C(t) = n^{1/2}(F_n(t) - F(t))$. We obtained a confidence interval for $F(t)$ at a fixed time t by obtaining values $L_{\alpha/2}(t), U_{\alpha/2}(t)$ such that $P\{L_{\alpha/2}(t) < n^{1/2}(F_n(t) - F(t)) < U_{\alpha/2}(t)\} = 1 - \alpha$.

D Let $R_n^D(p) = n^{1/2}(F_n^{-1}(p) - F^{-1}(p))$. We obtained a confidence interval for the p^{th} quantile of F by obtaining values $L_{\alpha/2}(p), U_{\alpha/2}(p)$ such that $P\{L_{\alpha/2}(p) < n^{1/2}(F_n^{-1}(p) - F^{-1}(p)) < U_{\alpha/2}(p)\} = 1 - \alpha$.

In study A we compared four estimates of the variance of the KME in terms of mean squared error in the random censorship model. Our simulations dealt with the case when the distributions were exponential. Before proceeding it is crucial to determine what is the true variance that we are trying to estimate. Let $\mathcal{C}$ be the information given in the data about the censoring variables. Since $\mathcal{C}$ is ancillary for the estimation of any characteristic of the survival curve, by the ancillary principle, the quantity that we should be estimating is $\mathrm{Var}(F_n(t) \mid \mathcal{C})$ and not $\mathrm{Var}(F_n(t))$. The exact conditional variance of $F_n(t)$ is not generally obtainable due to the dependence between the observed time T and indicator variable δ except for the case of the proportional hazards model (Chen, Hollander and Langberg, 1982). The exact *conditional* variance of $F_n(t)$ does not seem to be obtainable even for the proportional hazards model. We note that the conditional variance varies from sample to sample. To evaluate the mean squared error of the bootstrap estimates of the conditional variance for 10 samples, we numerically proceeded as follows.

Step 1 Get one set of data $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ by generating the failure time X from the standard exponential distribution and the censoring time Y from the exponential distribution with parameter λ. Here, λ was chosen so that $P(X > Y) = w$, where w was set at .2, .4 and .6.

Step 2 This step is for evaluating numerically $\mathrm{Var}(F_n(t) \mid \mathcal{C})$ where $\mathcal{C}$ is the censoring pattern observed in the data set $(T_1, \delta_1), \ldots, (T_n, \delta_n)$ in Step 1. First we generated one data set "compatible" with the censoring pattern $\mathcal{C}$ in the data set generated in Step 1, as follows. For $i = 1, \ldots, n$,

- generate $X_i' \sim$ Exponential(1);
- generate Y_i' conditionally on $\mathcal{C}$: If $\delta_i = 0$, then $Y_i' = T_i$. If $\delta_i = 1$, then generate Y_i' from the exponential distribution with parameter λ, conditional on being greater than T_i; note that by the memoryless property of the exponential distribution, this is equivalent to $Y_i' = T_i +$ Exponential(λ);
- form (T_i', δ_i') from X_i' and Y_i'.

Get the KME for the data set $(T_1', \delta_1'), \ldots, (T_n', \delta_n')$. We repeated this a large number of times (say, 100,000) and obtained numerically the true conditional variance of $F_n(t)$, $\mathrm{Var}(F_n(t) \mid \mathcal{C})$. Note that $\mathcal{C}$ remained fixed for the 100,000 repetitions.

Step 3 Generate one data set compatible with the censoring result in Step 1 as was done in Step 2. Calculate the four estimates of the variance of $F_n(t)$, one by Greenwood's formula and three by Bootstrap Methods 1, 2 and 3 (Bootstrap Methods 1 and 2 are equivalent but we applied both methods for the sake of assessing the random error). Repeat this step many (1,000) times and get the (estimated) mean squared error for the various estimates.

Step 4 Repeat Steps 1 through 3 ten times to observe the effect of various censoring patterns.

The results of this Monte Carlo study are shown in Table 2.1, where the mean squared errors for the four estimates are reported. In our study, we also obtained the biases and variances of the estimates, but these are not shown in the table. Table 2.1(a) pertains to the estimates of variance, whereas Table 2.1(b) pertains to the estimates of standard deviation. For Study A we make the following observations.

- For the variance estimates, when censoring was heavy, the estimate by Bootstrap Method 3 was more biased but much less variable, giving appreciably smaller mean squared error (the ratio of mean squared error becomes as large as 1.14) compared with Bootstrap Method 1 or 2.
- For the variance estimates, the estimate by Greenwood's formula performed substantially better than the estimates by any of the bootstrap methods except in the tail and when censoring was heavy. It gave larger mean squared error in the tail with heavy censoring because it was severely biased downward there. This phenomenon is well documented; see for example Peto et al. (1977).
- For the standard deviation estimates the results were very similar to those of the variance estimates except in the tail. There, the bias of the estimate by Method 3 was substantial enough so that the overall mean squared error for Method 3 was larger than for Methods 1 and 2.

An interesting conclusion to be drawn from this study is that for estimating the variance or the standard deviation of the KME, the estimate by Greenwood's formula should be preferred over the others (even though Bootstrap Method 3 performed better than other bootstrap methods). However, it should be kept in mind that our main objective here is to compare bootstrap methods, as these are far more general in their applicability. For example, there is no analogue to Greenwood's formula for estimating the variability of estimates of quantiles. (Actually one can estimate the variability of quantile estimates; however, this requires estimation of the density, which is difficult for small samples.)

For simulation study B, bootstrap confidence bands were obtained and compared in the random censorship model and in the model with two types of censoring. We used the bootstrap confidence bands derived from the following proposition.

Proposition 1. (Akritas, 1986). Suppose that F is nondegenerate. Choose $c_n(F_n)$ from the bootstrap distribution so that

$$P^* \left\{ n^{1/2} \sup_{0 \le t \le \tau} \left(\left| [\overline{F}^*_n(t) - \overline{F}_n(t)][\overline{B}_n(t)/\overline{F}_n(t)] \right| \right) \le c_n(F_n) \right\} = 1 - \alpha,$$

where

$$\overline{B}_n(t) = [1 + C_n(t)]^{-1}, \; C_n(t) = n \sum_{i:T_{(i)} \le t} \delta_i (n-i)^{-1} (n-i+1)^{-1}.$$

Then if $D_n(t) = n^{-1/2} \overline{F}_n(t)/\overline{B}_n(t)$,

$$P \left\{ \overline{F}_n(t) - c_n(F_n) D_n(t) \le \overline{F}(t) \le \overline{F}_n(t) + c_n(F_n) D_n(t), 0 \le t \le \tau \right\} \to 1 - \alpha.$$

Since all bootstrap methods have the same valid limiting distribution as shown in Theorems 1 and 2 in Section 3, we can apply the above proposition to all bootstrap methods. For constructing a $100(1-\alpha)\%$ bootstrap confidence band, we computed

$$n^{1/2} \sup \left\{ \left| [\overline{F}^*_n(t) - \overline{F}_n(t)] \frac{\overline{B}_n(t)}{\overline{F}_n(t)} \right| ; 0 \le t \le \tau \right\}$$

for each of 2,000 bootstrap samples and then approximated $c_n(F_n)$ by the $(1-\alpha)100^{\text{th}}$ percentile of these numbers. The number of simulations was 1,000 and the supremum was evaluated up to the second largest uncensored observation.

To evaluate the performance of the bands we considered two criteria, the coverage probability and the width of the bands. Before proceeding we need to explain exactly what we mean by "coverage probability" of a band $\{(L(t), U(t)) ; 0 \le t \le \tau\}$. Since, as was explained earlier, $\mathcal{C}$ is ancillary, we should condition on it and therefore we should consider $P\{L(t) \le F(t) \le U(t);\ 0 \le t \le \tau \mid \mathcal{C}\}$, the conditional coverage probability given $\mathcal{C}$. Thus, to estimate the conditional coverage probability given $\mathcal{C}$ of the band, we began by generating one data set and we obtained the censoring pattern for that data set. Then we generated 1,000 data sets compatible with that censoring pattern, computed the bands for each of these 1,000 data sets, and recorded the proportion of bands that contain the true distribution function. This whole process was repeated 10 times, giving results for 10 different censoring patterns (Tables 2.2 and 2.3 give the result for one censoring pattern; the results for other censoring patterns were similar). The description of this process for the case of two types of censoring is very similar, but nevertheless we describe it in detail below.

Step 1 For each $i = 1, \ldots, n$, the failure time X_i was generated from the standard exponential distribution, the censoring time C_i from the exponential distribution with parameter λ and the fixed censoring time w_i due to the termination of the study from the uniform distribution over $(0, b)$. Here, λ and b were chosen to get the proper censoring weight. For example $\lambda = .42013, b = 4.7604$ for 40% censoring (i.e. $P(X > Y) = .4$), and $\lambda = .90587, b = 2.2078$ for 60% censoring. Then form one data set (T_i, w_i, δ_i) by taking $T_i = \min(X_i, C_i, w_i)$, $\delta_i = 1$ if $T_i = X_i$, $\delta_i = .5$ if $T_i = C_i$ and $\delta_i = 0$ otherwise.

Step 2 To get a data set compatible with the censoring result in Step 1, generate $X_i' \sim$ Exponential(1), and if $\delta_i = .5$, set $C_i' = T_i$; otherwise set $C_i' = T_i +$ Exponential(λ). Then form (T_i', w_i', δ_i'), where $w_i' = w_i$, $T_i' = \min(X_i', C_i', w_i')$, and $\delta_i' = 1$ if $T_i' = X_i'$, $\delta_i' = .5$ if $T_i' = C_i'$, and $\delta_i' = 0$ otherwise.

Step 3 From the compatible data set generated in Step 2, take bootstrap samples and apply Bootstrap Method 3′ to get the bootstrap confidence bands.

Step 4 Repeat Steps 2 and 3 many (1,000) times. The proportion of times that the band contained the true distribution function was taken as the estimate of the conditional coverage probability given the censoring pattern.

Table 2.2 summarizes the results of Study B for the model with two types of censoring — fixed time censoring due to the termination of the study (uniform distribution) and censoring due to other causes (exponential distribution). In the table, $c_n(F_n)$ reflects essentially the width of confidence bands, since $D_n(t)$ varies with time t but remains the same for all bootstrap methods. Table 2.2 shows that in heavy censoring we got a very minor gain in the width of the confidence band by Method 3′. For Table 2.3 we assumed the standard random censorship model where the censoring times $Y_1, \ldots, Y_n$ were iid as in Simulation Study A. We note that Bootstrap Method 3′ becomes identical to Method 3 if all fixed censoring times are infinite or are greater than any censoring time due to unrelated causes. But for the sake of assessing the random error we applied Method 3′ here for Table 2.3. Method 3 bands had slightly narrower width than Method 1 (and 2) bands. For example, in the case of 60% censoring, the ratio of average widths of 95% confidence bands was 1.019. Note that the results of Method 1 and Method 2 were very close, and so were those of Method 3 and Method 3′, as expected. This led us to believe that although the advantage gained by using Method 3 was minor, it was statistically significant, i.e. it was not due to random fluctuation.

Simulation Study C and D were done, but not reported here. The reader is suggested to see Kim (1990) for detailed description and result.

The overall conclusions of the simulation studies are that Bootstrap Methods 1 (or 2) became more different from 3 as the censoring weight got heavier, but that there did not appear to be a practical difference between those methods. It should be kept in mind, however, that Method 3 can be easily extended to Method 3′ which is the only method with a firm statistical basis in the more general setup in which there are two types of censoring.

Table 2.1. (a) Mean squared errors of variance estimates of KME. The table gives combined result for ten censoring patterns.
(number of bootstraps=800, number of simulations=1,000. The entry in parentheses is the estimated standard error.)

n	cw	t	MSEG	MSE1	MSE2	MSE3	Ratiol	Ratio2
30	20%	0.29	.0019	.0020	.0020	.0020	1.004 (.016)	0.950 (.015)
		0.51	.0007	.0009	.0009	.0009	1.011 (.039)	0.815 (.018)
		0.69	.0003	.0005	.0005	.0005	1.000 (.044)	0.639 (.038)
		0.92	.0008	.0010	.0010	.0010	0.987 (.021)	0.812 (.023)
		1.39	.0032	.0034	.0034	.0033	1.018 (.013)	0.959 (.021)
	40%	0.29	.0028	.0030	.0030	.0029	1.018 (.013)	0.959 (.010)
		0.51	.0016	.0020	.0019	.0019	1.021 (.017)	0.854 (.022)
		0.69	.0015	.0019	.0019	.0019	1.032 (.034)	0.817 (.041)
		0.92	.0034	.0040	.0039	.0037	1.049 (.024)	0.916 (.036)
		1.39	.0199	.0199	.0199	.0186	1.067 (.012)	1.043 (.067)
	60%	0.29	.0042	.0045	.0045	.0044	1.024 (.017)	0.959 (.014)
		0.51	.0072	.0092	.0091	.0082	1.092 (.048)	0.877 (.017)
		0.69	.0140	.0188	.0188	.0162	1.141 (.052)	0.864 (.023)
		0.92	.0422	.0470	.0470	.0424	1.115 (.030)	0.978 (.056)
		1.39	.2479	.1903	.1901	.1723	1.102 (.042)	1.402 (.094)
40	20%	0.29	.0009	.0009	.0009	.0009	1.005 (.015)	0.941 (.011)
		0.51	.0003	.0004	.0004	.0004	1.020 (.031)	0.766 (.021)
		0.69	.0001	.0003	.0003	.0003	1.001 (.045)	0.556 (.054)
		0.92	.0004	.0005	.0005	.0005	0.992 (.031)	0.771 (.022)
		1.39	.0015	.0016	.0016	.0015	1.010 (.015)	0.949 (.014)
	40%	0.29	.0011	.0012	.0012	.0012	1.017 (.020)	0.949 (.016)
		0.51	.0007	.0009	.0009	.0008	1.042 (.031)	0.836 (.031)
		0.69	.0006	.0008	.0008	.0008	1.021 (.026)	0.743 (.056)
		0.92	.0013	.0015	.0015	.0015	1.029 (.031)	0.854 (.028)
		1.39	.0078	.0083	.0084	.0080	1.047 (.021)	0.971 (.027)
	60%	0.29	.0019	.0020	.0020	.0020	1.030 (.020)	0.955 (.012)
		0.51	.0025	.0030	.0030	.0028	1.055 (.035)	0.890 (.021)
		0.69	.0052	.0069	.0068	.0061	1.125 (.037)	0.856 (.030)
		0.92	.0176	.0219	.0221	.0195	1.124 (.023)	0.896 (.044)
		1.39	.1177	.0908	.0914	.0879	1.037 (.010)	1.305 (.115)

Key to Table 2.1 (a)

n=sample size: cw=censoring weight.

t=time, corresponding to the .25, .4, .5, .6 and .75$^{\text{th}}$ quantiles of the standard exponential distribution function.

MSEG, MSE1, MSE2, MSE3=average of ten observed mean squared error($\times 1,000$) of the variance estimate of the KME over ten censoring patterns by Greenwood's formula and Bootstrap Methods 1, 2 and 3, respectively.

Ratiol=average of ten ratios of the observed mean squared error by Method 1 (and 2) to the observed mean squared error by Method 3.

Ratio2=average of ten ratios of the observed mean squared error by Greenwood's formula to the observed mean squared error by Method 3.

The estimated standard error of each ratio (Ratiol and Ratio2) was calculated using the formula: standard error estimate = $\sqrt{\sum_{i=1}^{10}(R_i - \bar{R})^2/9}$ where R_i is the observed ratio for the i^{th} censoring pattern.

Table 2.1. (b) Mean squared errors of standard deviation estimates of KME. The table gives combined result for ten censoring patterns. (number of bootstraps=800, number of simulations=1,000. The entry in parentheses is the estimated standard error.)

n	cw	t	MSEG	MSE1	MSE2	MSE3	Ratio1	Ratio2
30	20%	0.29	0.0911	0.0953	0.0949	0.0948	1.004 (.013)	0.961 (.013)
		0.51	0.0239	0.0290	0.0286	0.0286	1.007 (.035)	0.833 (.016)
		0.69	0.0102	0.0158	0.0154	0.0157	0.996 (.044)	0.651 (.040)
		0.92	0.0250	0.0295	0.0297	0.0301	0.983 (.022)	0.828 (.021)
		1.39	0.1369	0.1325	0.1327	0.1315	1.007 (.011)	1.034 (.056)
	40%	0.29	0.1203	0.1259	0.1251	0.1246	1.008 (.011)	0.966 (.008)
		0.51	0.0473	0.0547	0.0541	0.0540	1.009 (.020)	0.872 (.025)
		0.69	0.0372	0.0449	0.0443	0.0438	1.014 (.033)	0.840 (.050)
		0.92	0.0816	0.0840	0.0832	0.0818	1.021 (.020)	0.993 (.083)
		1.39	0.7403	0.4539	0.4546	0.4563	0.998 (.012)	1.550 (.272)
	60%	0.29	0.1706	0.1772	0.1775	0.1760	1.007 (.013)	0.968 (.012)
		0.51	0.1561	0.1818	0.1816	0.1719	1.048 (.030)	0.905 (.017)
		0.69	0.2403	0.2717	0.2713	0.2516	1.070 (.027)	0.939 (.055)
		0.92	0.7697	0.5559	0.5584	0.5564	1.012 (.038)	1.332 (.176)
		1.39	5.5550	1.9104	1.9080	2.0820	0.921 (.029)	2.629 (.237)
40	20%	0.29	0.0513	0.0540	0.0544	0.0540	1.003 (.012)	0.949 (.010)
		0.51	0.0134	0.0173	0.0175	0.0172	1.014 (.028)	0.780 (.021)
		0.69	0.0055	0.0094	0.0096	0.0096	0.997 (.044)	0.564 (.054)
		0.92	0.0143	0.0178	0.0179	0.0182	0.986 (.030)	0.784 (.021)
		1.39	0.0705	0.0726	0.0724	0.0723	1.002 (.014)	0.973 (.019)
	40%	0.29	0.0638	0.0673	0.0675	0.0667	1.012 (.020)	0.957 (.016)
		0.51	0.0264	0.0316	0.0322	0.0309	1.034 (.029)	0.850 (.032)
		0.69	0.0179	0.0237	0.0232	0.0232	1.009 (.027)	0.753 (.059)
		0.92	0.0368	0.0419	0.0418	0.0415	1.010 (.030)	0.875 (.035)
		1.39	0.3051	0.2290	0.2312	0.2315	1.001 (.018)	1.251 (.189)
	60%	0.29	0.0919	0.0973	0.0966	0.0952	1.018 (.017)	0.964 (.012)
		0.51	0.0733	0.0835	0.0829	0.0804	1.030 (.027)	0.909 (.022)
		0.69	0.1121	0.1341	0.1334	0.1247	1.073 (.028)	0.899 (.051)
		0.92	0.3624	0.3281	0.3298	0.3149	1.048 (.021)	1.133 (.120)
		1.39	3.2827	1.2265	1.2283	1.3375	0.926 (.031)	2.395 (.324)

Key to Table 2.1 (b)

n=sample size: cw=censoring weight.

t=time, corresponding to the .25, .4, .5, .6 and .75$^{\text{th}}$ quantiles of the standard exponential distribution function.

MSEG, MSE1, MSE2, MSE3=average of ten observed mean squared error($\times 1,000$) of the standard deviation estimate of the KME over ten censoring patterns by Greenwood's formula and Bootstrap Methods 1, 2 and 3, respectively.

Ratio1=average of ten ratios of the observed mean squared error by Method 1 (and 2) to the observed mean squared error by Method 3.

Ratio2=average of ten ratios of the observed mean squared error by Greenwood's formula to the observed mean squared error by Method 3.

The estimated standard error of each ratio (Ratio1 and Ratio2) was calculated using the formula: standard error estimate = $\sqrt{\sum_{i=1}^{10}(R_i - \bar{R})^2/9}$ where R_i is the observed ratio for the i^{th} censoring pattern.

Table 2.2. Bootstrap confidence bands of survival curve for model with two types of censoring. The table gives the result for one censoring pattern. (number of bootstraps=2,000, number of simulations=1,000 and sample size n=40. The entry in parentheses is the estimated standard error.)

		Method 1		Method 2		Method 3′		
cw	1-α	C.P.	$c_n(F_n)$	C.P.	$c_n(F_n)$	C.P.	$c_n(F_n)$	Ratio
40%	.95	.964	1.241	.962	1.242	.959	1.232	1.008
		(.006)	(.001)	(.006)	(.001)	(.006)	(.001)	(.001)
	.90	.926	1.108	.922	1.109	.920	1.100	1.007
		(.008)	(.000)	(.008)	(.000)	(.009)	(.001)	(.001)
	.80	.826	0.958	.823	0.958	.823	0.955	1.004
		(.012)	(.000)	(.012)	(.000)	(.012)	(.001)	(.001)
60%	.95	.984	1.209	.981	1.209	.977	1.194	1.013
		(.004)	(.001)	(.004)	(.001)	(.005)	(.001)	(.001)
	.90	.953	1.070	.952	1.070	.950	1.060	1.010
		(.007)	(.001)	(.007)	(.001)	(.007)	(.001)	(.001)
	.80	.878	0.916	.880	0.915	.874	0.908	1.009
		(.010)	(.001)	(.010)	(.001)	(.010)	(.001)	(.001)

Table 2.3. Bootstrap confidence bands of survival curve for standard random censorship model. The table gives the result for one censoring pattern. (number of bootstraps=2,000, number of simulations=1,000 and n=40. The entry in parentheses is the estimated standard error.)

		Method 1		Method 2		Method 3		Method 3′		
cw	1-α	C.P.	$c_n(F_n)$	C.P.	$c_n(F_n)$	C.P.	$c_n(F_n)$	C.P.	$c_n(F_n)$	Ratio
40%	.95	.962	1.238	.962	1.240	.962	1.229	.958	1.229	1.008
		(.006)	(.001)	(.006)	(.001)	(.006)	(.001)	(.006)	(.001)	(.001)
	.90	.922	1.103	.922	1.104	.918	1.094	.914	1.095	1.008
		(.008)	(.001)	(.008)	(.001)	(.009)	(.001)	(.009)	(.001)	(.001)
	.80	.837	0.952	.840	0.954	.834	0.946	.832	0.947	1.006
		(.012)	(.000)	(.012)	(.000)	(.012)	(.001)	(.012)	(.001)	(.001)
60%	.95	.980	1.197	.981	1.195	.979	1.175	.979	1.175	1.019
		(.004)	(.001)	(.004)	(.001)	(.005)	(.001)	(.005)	(.001)	(.001)
	.90	.944	1.055	.946	1.054	.942	1.039	.941	1.039	1.016
		(.007)	(.001)	(.007)	(.001)	(.007)	(.001)	(.007)	(.001)	(.001)
	.80	.877	0.899	.876	0.899	.872	0.888	.873	0.888	1.013
		(.010)	(.001)	(.010)	(.001)	(.011)	(.001)	(.011)	(.001)	(.001)

Key to Tables 2.2 and 2.3:

cw=censoring weight.

C.P.=observed coverage probability of confidence bands.

$c_n(F_n)$=average of 1,000 $c_n(F_n)$ in Proposition 1 which is proportional to the width of confidence band.

Ratio=in Table 2.2 average of 1,000 ratios of width 1 to width 3′, in Table 2.3 average of 1,000 ratios of width 1 to width 3.

The estimated standard error of Ratio (average ratio) was obtained from 1,000 ratios.

3. PROOFS OF WEAK CONVERGENCE RESULTS

Akritas (1986) proved weak convergence of the bootstrapped KME for Method 1. In this section the weak convergence of a properly standardized KME by Bootstrap Methods 3 and 3′ is proved. Theorem 1 pertains to the standard random censorship model and Theorem 2 pertains to the model with two types of censoring. It turns out that the properly normalized bootstrapped versions of the KME by Methods 1, 3, and 3′ have the same limiting distributions.

Theorem 1. Assume the random censorship model where the failure time X has distribution function F and the censoring time Y has continuous distribution function G. Let H be the distribution function of the observed time ($H = 1-(1-F)(1-G)$). Let F_n^* be the Kaplan-Meier estimate of F_n computed from the data resampled by Bootstrap Method 3. Then for almost every infinite sequence $(T_1,\delta_1),(T_2,\delta_2),\ldots$, as $n \to \infty$

$$\sqrt{n}\left(\frac{F_n^* - F_n}{1- F_n}\right) \xrightarrow{d} W \text{ in } D[0,\tau]$$

for any $\tau < T = \sup\{t : H(t) < 1\}$, where W is a mean zero Gaussian process with independent increments and variance function given by

$$\text{Var}\,(W(t)) = \int_{[0,t]} \frac{1}{(1-F(s))(1-H_-(s))}\, dF(s).$$

Proof. For Bootstrap Method 3 we have the *General Random Censorship Model* (see Gill 1980, p. 55) since each censoring variable has a different distribution. Theorem 4.2.1 of Gill (1980) is general enough to accommodate:

- the General Random Censorship Model.
- the dependence of F_n, which is the distribution function of the resampled failure time X_j^*, on n.
- the discontinuity of F_n.

The martingale arguments of Gill (1980) apply conditionally on a specified sequence $\{(T_i,\delta_i), i = 1,\ldots\}$. The proof of our weak convergence theorem consists mainly of two steps:

For almost every infinite sequence $\{(T_i, \delta_i), i = 1, \ldots\}$ (we will use the terminology *conditionally almost surely*),

1. $Z_n^*(t) = \sqrt{n}\left(\frac{F_n^*(t) - F_n(t)}{1 - F_n(t)}\right), t \in [0, \tau]$ can be represented as a stochastic integral with respect to a martingale and is therefore a martingale (for a more precise statement, see (3.2), (3.3) and the remarks following (3.4) below).
2. The martingale Z_n^* satisfies the Conditions (I) of Theorem 4.2.1 of Gill (1980) which are sufficient conditions for the weak convergence of the martingale Z_n^*.

For Bootstrap Method 3, we resample

$$X_j^* \sim F_n,$$

$$Y_j^* \sim L_j^n \text{ which is represented as}$$
$$L_j^n(t) = I(\delta_j = 0) I(t \in [T_j, \infty)) + I(\delta_j = 1)\frac{(G_n(t) - G_n(T_j) I(t \in [T_j, \infty))}{1 - G_n(T_j)},$$

then form $T_j^* = \min(X_j^*, Y_j^*)$, $\delta_j^* = I(X_j^* \leq Y_j^*)$; $j = 1, \ldots, n$. (Note: In his treatment of the General Random Censorship Model, Gill allows the possibility that L_j^n be a subdistribution function.)

We define stochastic processes by

$$N_n^*(t) = \#\{j : T_j^* \leq t \text{ and } \delta_j^* = 1, j = 1, \ldots, n\},$$
where # denotes the number of elements in a set,
$$V_n^*(t) = \#\{j : T_j^* \geq t\},$$
$$M_n^*(t) = N_n^*(t) - \int_{[0,t]} V_n^*(s)\, d\Lambda_n(s),$$
where the function Λ_n is defined by
$$\Lambda_n(t) = \int_{[0,t]} \frac{1}{1 - F_{n-}(s)}\, dF_n(s),$$
$$J_n^*(t) = I(V_n^*(t) > 0).$$

It can be shown that M_n^* is a zero-mean square-integrable martingale with bracket function

$$\langle M_n^* \rangle(t) = \int_{[0,t]} V_n^*(s)(1 - \Delta\Lambda_n(s))\, d\Lambda_n(s), \tag{3.1}$$

where $\Delta A = A - A_-$ for any right continuous function with left-hand limits (see Appendix A of Kim (1990) for the proof that M_n^* is a martingale with respect to an appropriate filtration).

Define the stopped process $F_n^\dagger$ on $[0, \infty)$ by

$$F_n^\dagger(t) = F_n(t \wedge T_{(n)}^*) \tag{3.2}$$

and let

$$Q_n^*(t) = \sqrt{n}\left(\frac{F_n^*(t) - F_n^\dagger(t)}{1 - F_n^\dagger(t)}\right) \text{ for } t \in [0, \infty). \tag{3.3}$$

Let τ_n be any number such that $F_n(\tau_n) < 1$. Then from equation (3.2.13) of Gill (1980) we have the representation

$$Q_n^*(t) = \sqrt{n}\int_{[0,t]} \frac{1 - F_{n-}^*(s)}{1 - F_n(s)} \frac{J_n^*(s)}{V_n^*(s)}\, dM_n^*(s) \text{ for } t \in [0, \tau_n]. \tag{3.4}$$

We remark that $Z_n^*(t) = \sqrt{n}(\frac{F_n^*(t) - F_n(t)}{1 - F_n(t)})$ need not be a martingale since the expected value of this quantity is 0 for $t = 0$ but need not be 0 for $t > 0$. However

$$\begin{aligned} P^*\{\sup_{0 \le t \le \tau} | Q_n^*(t) - Z_n^*(t) | \ne 0\} &= P^*\{T_{(n)}^*(t) < \tau\} \\ &\le \{H_n(\tau)\}^n \to 0 \text{ conditionally a.s.} \end{aligned}$$

since $H_n(\tau) \to H(\tau) < 1$ a.s. Hence $Q_n^*(t)$ and $Z_n^*(t)$ have the same limiting distribution. Moreover, since a.s. as $n \to \infty$ $F_n(\tau) \to F(\tau) < 1$, the representation (3.4) is valid over $[0, \tau]$ conditionally a.s. for large n. Finally since $J_n^*(s) = I(0 \le s \le T_{(n)}^*)$, $J_n^*(s)$ can be trivially ignored in the representation (3.4).

Relations (3.1) and (3.4) imply that Q_n^* is a square-integrable martingale with bracket function

$$\langle Q_n^* \rangle(t) = \int_{[0,t]} \left(\frac{1 - F_{n-}^*(s)}{1 - F_n(s)}\right)^2 \frac{n J_n^*(s)}{V_n^*(s)}(1 - \Delta\Lambda_n(s))\, d\Lambda_n(s).$$

Now we proceed to check Conditions (I) of Theorem 4.2.1 of Gill (1980). The bootstrap version of Conditions (I) is as follows.

Conditions (I):

a. F_n converges uniformly on $[0, \tau]$ a.s. to F as $n \to \infty$; $\Lambda = \int \frac{1}{1-F_-} dF$ is finite on $[0, \tau]$.

b. There is a function h that is left continuous with right-hand limits and is of bounded variation on $[0, \tau]$ such that $\left(\frac{1-F^*_{n-}}{1-F_n}\right)^2 \frac{n}{V^*_n} J^*_n$ converges uniformly on $[0, \tau]$ in conditional probability a.s. to h.

c. $V^*_n(t) \to \infty$ in conditional probability a.s. as $n \to \infty$ for each $t \in [0, \tau]$.

Conditions (I).a is clear by the strong uniform consistency of F_n (Földes, Rejtö and Winter, 1980) and the definition of τ. To check Condition (I).b we need the following lemma. The proof of the lemma can be found in Kim (1990).

Lemma 1. When bootstrapping is done by Method 3,

$$\frac{1}{n} \sum_{j=1}^{n} L^n_j(t) \to G(t) \text{ uniformly on } [0, \tau] \text{ a.s. as } n \to \infty.$$

As pointed out in Gill (1980, p. 70) it can be shown that Lemma 1 is a sufficient condition for

$$\frac{V^*_n}{n} \xrightarrow{\mathcal{P}^*} (1 - F_-)(1 - G_-) = 1 - H_- \text{ uniformly on } [0, \tau] \text{ a.s.}. \quad (3.5)$$

Moreover, since $\sup_{0 \le t \le \tau} \mid F^*_n(t) - F_n(t) \mid \xrightarrow{\mathcal{P}^*} 0$ (Theorem 4.1.1 of Gill, 1980) and $\sup_{0 \le t \le \tau} \mid F_n(t) - F(t) \mid \to 0$ a.s., we have

$$F^*_{n-}(t) \xrightarrow{\mathcal{P}^*} F_-(t) \text{ for all } t \text{ in } [0, \tau]. \quad (3.6)$$

Let $h = \left(\frac{1-F_-}{1-F}\right)^2 \frac{1}{1-H_-}$. From (3.5) and (3.6), we have

$$\left(\frac{1 - F^*_{n-}}{1 - F_n}\right)^2 \frac{n}{V^*_n} \xrightarrow{\mathcal{P}^*} h \quad \text{conditionally a.s..}$$

Assuming now that F is continuous, then h is left continuous with right-hand limits and $h+$ is of bounded variation, so that Condition (I).b is satisfied. Relation (3.5) implies Condition (I).c. Now since all parts of Conditions (I) are satisfied, from Theorem 4.2.1 of Gill (1980) and the remarks following (3.4) we have

$$Z_n^* \xrightarrow{d} W \text{ in } D[0,\tau] \text{ conditionally a.s.,}$$

where the variance function of the limiting mean zero Gaussian process with independent increments is

$$\int h(1-\Delta\Lambda)\, d\Lambda = \int \frac{1}{(1-F)(1-H_-)}\, dF$$

(note that $1-F=(1-\Delta\Lambda)(1-F_-)$).

For the case in which F has discontinuities, the same result can be proved by the arguments of Akritas (1986, p. 1037). In those arguments, the general random censorship model in resampling has no role, so that the arguments can be applied directly to our situation. Now the proof of the theorem is complete. □

In Theorem 1 we assume that the censoring distribution is continuous. The assumption of continuity of the censoring distribution simplifies the proof of uniform convergence in Lemma 1. For Bootstrap Method 3′ in the model with two types of censoring, the same weak convergence result is proved. Once again we assume the continuity of the censoring distributions. We conjecture that the same results hold for any censoring distributions.

Theorem 2. Suppose that the failure time X has distribution function F and the censoring time C due to other causes has a continuous distribution function K. Assume that all the fixed censoring times $w_1,\ldots,w_n$ are known and that the empirical c.d.f. of the w's converges to a continuous c.d.f. U for all t. Suppose that the failure time, the fixed censoring time and the censoring time due to other causes are independent. Let $H=1-(1-F)(1-K)(1-U)=1-(1-F)(1-G)$. Let F_n^* and K_n^* be the Kaplan-Meier estimates of F_n and K_n, computed from the data resampled by Bootstrap Method 3′, respectively. Then for almost every infinite sequence $(T_1,w_1,\delta_1),(T_2,w_2,\delta_2),\ldots$, as $n\to\infty$

$$\sqrt{n}\left(\frac{F_n^*-F_n}{1-F_n}\right) \xrightarrow{d} W \text{ in } D[0,\tau]$$

for any $\tau<T=\sup\{t: H(t)<1\}$, where W is a mean zero Gaussian process with independent increments and variance function given by

$$\mathrm{Var}\,(W(t)) = \int_{[0,t]} \frac{1}{(1-F(s))(1-H_{-}(s))}\, dF(s).$$

Proof. For Bootstrap Method 3′, we resample

$$\begin{aligned} X_j^* &\sim F_n, \\ Y_j^* &\sim L_j^n \quad \text{which can be represented as} \\ & L_j^n(t) = I(\delta_j \neq 1) I(t \in [T_j, \infty)) \\ & \qquad + I(\delta_j = 1) I(t \in [T_j, w_j)) \frac{K_n(t) - K_n(T_j)}{1 - K_n(T_j)} \\ & \qquad + I(\delta_j = 1) I(t \in [w_j, \infty)), \end{aligned}$$

then form $T_j^* = \min(\, X_j^*,\ Y_j^*)$, $\delta_j^* = I(T_j^* = X_j^*) + .5 I(T_j^* < X_j^*, T_j^* < w_j)$; $j = 1, \ldots, n$. Theorem 2 follows from the lemma below. The proof of the lemma can be found in Kim (1990).

Lemma 2. When bootstrapping is done by Method 3′,

$$\frac{1}{n} \sum_{j=1}^{n} L_j^n(t) \to G(t) \text{ uniformly on } [0, \tau] \text{ as. as } n \to \infty. \ \Box$$

4. AN APPLICATION TO REAL DATA

In this section the various bootstrap methods are applied to a data set involving survival times after treatment for patients with malignant melanoma. In the period of 1964–73, 225 patients with malignant melanoma (cancer of the skin) had radical surgery performed at the Department of Plastic Surgery, University Hospital of Odense, Denmark. All patients were followed until the end of 1977. The survival time since the operation was censored by death from other causes and also by the termination of the follow-up period. Thus, associated with each individual is the vector (X_i, w_i, C_i) where X_i is the survival time if there was no censoring, w_i is the censoring time due to the termination of the study, and C_i is the time of death or withdrawal due to other causes. Note that the w_i's are known for *all* patients since we know the times of operation for all patients. A full listing of the data is given in Andersen, Borgan, Gill and Keiding (1993). In analyzing this data set, it was found that sex was a significant risk factor, with men having higher risk than women; see Andersen et al. (1993). As explained in Section 1, Bootstrap Method 3′ is the appropriate bootstrap method for this data set since we have two types of censoring, one due to the termination of study and the other due to other causes.

As pointed out in Akritas (1986), the bootstrap bands are not monotonic towards the tail; this was corrected by replacing the values of a band at these points of nonmonotonicity by the preceding values. We bootstrapped 4,000 times for both men and women. When we increased the number of bootstraps to 8,000 and 16,000 the resulting difference was negligible. Figure 1 shows that Method 3′ gives slightly narrower bands in the men's data.

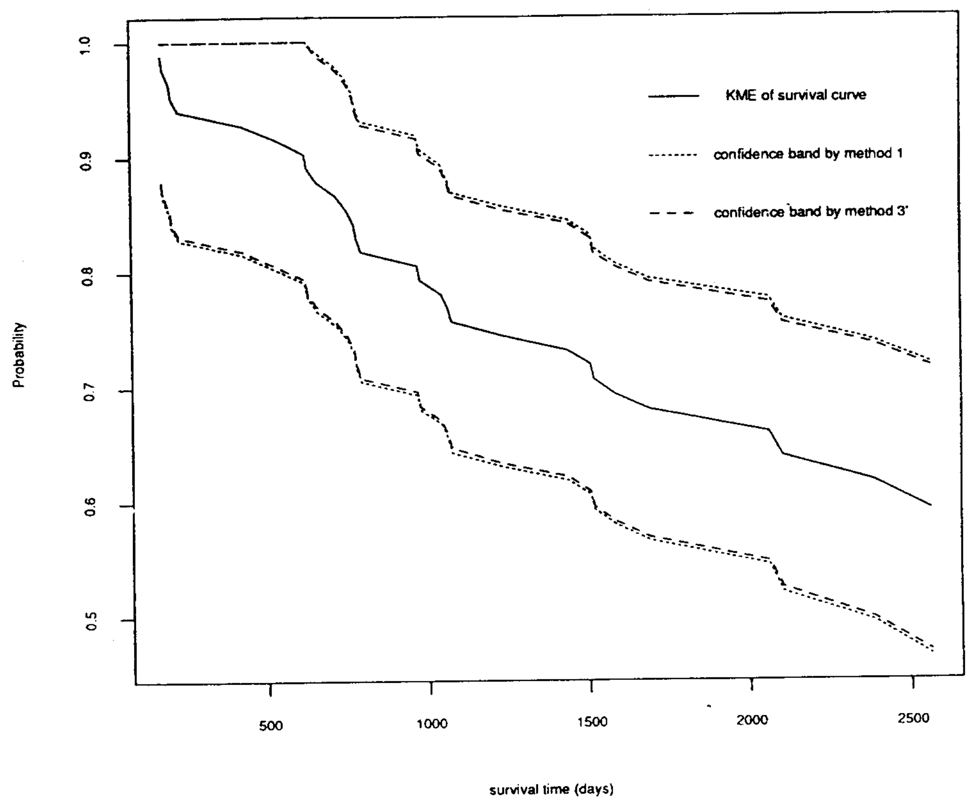

Figure 1. 90% boostrap confidence band of survival curve for men in melanoma data. (All the graphs were interpolated linearly between observations for convenience in plotting)

We also obtained bootstrap confidence intervals for survival curves at fixed times, and bootstrap confidence intervals for the $.25^{th}$ quantile of F in the data for the men. The results are reported in Kim (1990).

5. CONCLUSION

We proposed a new bootstrap scheme, called Method 3, that acts conditionally on the censoring pattern. The numerical results in Section 2 suggest that Method 3 performs at least as well as Efron's scheme (Method 1 and 2) overall, and gives minor gains in efficiency (i.e. smaller mean squared error for the variance estimate of the KME and narrower width of confidence bands or intervals) when censoring is heavy. Even though the gains obtained by Method 3 are minor, we believe that Method 3 should be preferred to Method 1 (2) since its statistical basis in the random censorship model is more sound. Moreover it can be extended to Method 3′ which is the most appropriate when we have the two types of censoring as in melanoma data described in Section 4.

REFERENCES

(1) Akritas, M.G. (1986). Bootstrapping the Kaplan-Meier estimator. *Journal of the American Statistical Association*, **81**, 1032–1038.

(2) Andersen, P.K., Borgan, Ø, Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

(3) Chung, K.L. (1974). *A Course in Probability Theory*, Second Edition. Academic Press, New York.

(4) Chen, Y.Y., Hollander, M. and Langberg, N.A. (1982). Small sample results for the Kaplan-Meier estimator. *Journal of the American Statistical Association*, **77**, 141–144.

(5) Efron, B. (1981). Censored Data and the Bootstrap. *Journal of the American Statistical Association*, **76**, 312–319.

(6) Efron, B. and Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, **1**, 54–75.

(7) Földes, A., Rejtö, L. and Winter, B. B. (1980). Strong consistency properties of nonparametric estimators for randomly censored data (Part I). *Periodica Mathematica Hungarica*. **11**, 233–250.

(8) Gill, R.D. (1980). *Censoring and Stochastic Integrals.* Mathematical Center Tracts **124**, Mathematisch Centrum Amsterdam.

(9) Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

(10) Kim, J (1990). *Conditional Bootstrap Methods.* Ph.D. Thesis, Florida State University.

(11) Liptser, R.S. and Shiryayev, A.N. (1978). *Statistics of Random Processes II : Applications.* Springer-Verlag, New York.

(12) Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V., Mantel, N., Mcpherson, K., Peto, J. and Smith, P.G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. *British Journal of Cancer*, **35**, 1–39.