

Some Applications of the *Bootstrap* to Survival Analysis

Frederic Utzet
Álex Sánchez
Universitat de Barcelona

En este trabajo se revisan algunas de las aplicaciones clásicas del bootstrap al análisis de la supervivencia. Se consideran en primer lugar el estimador bootstrap de la varianza y el estimador de la mediana corregido para el sesgo del estimador de Kaplan-Meier de la función de supervivencia. A continuación se consideran algunos aspectos más recientes, tales como métodos para construir bandas de confianza para el estimador de la función de supervivencia y contrastes aproximados para la comparación de funciones de supervivencia. En ambas situaciones el bootstrap resulta de gran utilidad para la aproximación de los valores críticos necesarios.

Palabras clave: Análisis de la supervivencia, datos censurados, estimador de Kaplan-Meier, bootstrap, contrastes de hipótesis bootstrap.

In this work some of the classical applications of bootstrap methods to survival analysis are reviewed. First the bootstrap estimate of the variance and the bias-corrected estimate of the median for the Kaplan-Meier estimator of the survival function are considered. After this more recent topics are considered such as methods for constructing confidence bands for the estimate of the survival function and exact and approximated tests to compare two survival functions. In the last two situations the bootstrap proves to be helpful in approximating the necessary critical values.

Key words: Survival Analysis, Censored Sample, Kaplan-Meier estimator, Bootstrap, Bootstrap test.

The Kaplan-Meier estimator (Kaplan and Meier, 1958) has a central role in the non-parametric approach to survival. In 1967 Efron gave a new insight to this estimator in an important paper (Efron, 1967), and therefore, it is not surprising that one of the first applications of the bootstrap was made in the con-

text of survival analysis (Efron, 1981). Here, in what follows we study two well known bootstrap estimators, both introduced by Efron (1981).

We begin with the *bootstrap estimate of the variance* of the Kaplan-Meier estimator at a point, which we compare with the classical asymptotic estimation given by the Greenwood formula. After this we consider the *bootstrap estimate of the bias* of the median of the distribution (also obtained using the Kaplan-Meier estimator).

Finally, a third application will be presented which deals with the crucial problem in survival analysis of testing whether two survival curves are the same. For example, suppose we have two (censored) samples recording the time of remission of a certain illness after the administration of 2 different drugs and we want to compare the efficiency of the two treatments. This is a version of the classical two-sample problem for censored samples. Here, one can adopt a parametric, semi-parametric or non-parametric point of view according to the hypotheses which are plausible in the model.

The approach we use in this paper is the non-parametric one. Some different, well-known tests, are available for this approach, known in general as the Mantel-Gehan-Efron-Cox tests. They are based on the weighted log-rank statistics (Fleming and Harrington, 1991), and can be rewritten in a general formulation (Gill, 1980) as a class of statistics based on the Kaplan-Meier estimator, which Gill calls *calls-K* statistics.

Any of these tests can be bootstrapped, but here we adopt a different point of view and we will study a Kolmogorov-Smirnov type test, which is intuitive in this situation and easy to bootstrap. Essentially, what we are going to do is to change the empirical distribution functions by the Kaplan-Meier estimations. If we assume that the (common) survival function is continuous, the statistic we will use is (almost) independent of the particular distribution, and we can construct an asymptotic test. This class of tests has been studied by Schumacher (1984). When no assumption about continuity is made, we can resort to a bootstrap test. The bootstrap works well in this situation, Efron (1981), Akritas (1986), Gill (1989), at least as well as the classical asymptotic theory, and it could work even better.

Notation. The Kaplan-Meier Estimator

Let X be a non-negative random variable. We are interested in the estimation of the survival function:

$$S_X(t) = P(X > t)$$

When we take a sample

$$X_1, \dots, X_n$$

it is not always possible to observe X ; sometimes we observe another random variable D . For example, say X is the time until remission of an illness under a cer-

tain treatment, but patient number i changes doctor before he is cured and is lost for our study. The general formulation is the following: let

$$(X_1, D_1), \dots, (X_n, D_n)$$

be i.i.d. vectors, X and D independents. Assume we observe

$$U_i = \min(X_i, D_i)$$

(D are called the censors), and suppose we also know whether or not the observation U is censored. This is given by a random variable

$$\delta_i = \begin{cases} 1, & \text{if } U_i = X_i \text{ (uncensored observation)} \\ 0, & \text{if } U_i = D_i \text{ (censored observation)} \end{cases}$$

To estimate the survival function using all the information, i.e.,

$$(U_1, \delta_1), \dots, (U_n, \delta_n)$$

we introduce the Kaplan-Meier estimator. Assume the observations are ordered:

$$U_1 < U_2 < \dots < U_n$$

then,

$$\hat{S}_X^n(t) = \prod_{U_j \leq t} \left(\frac{n-j}{n-j+1} \right)^{\delta_j}, \text{ if } t \leq U_n$$

and

$$\hat{S}_X^n(t) = 0, \text{ if } t > U_n$$

When there is no censored observation the Kaplan-Meier estimator is just the empirical survival function (which assigns mass $1/n$ to each observation).

EXAMPLE 1¹

Most examples that will be presented here are based on the data (so many times analyzed and reanalyzed) from Freireich *et al.* (1963) (see for example Lawless, 1982) in which a drug was compared to a placebo in the capacity to maintain remission in acute leukaemia patients:

The data are the following: (*denotes a censored observation)

1. All the examples have been calculated using OBJECSTAT and EMSS which are a set of object-oriented statistical routines and an integrated environment for its use developed by one of the authors (see Sánchez, Ocaña and Ruiz de Villa, 1992).

6-MP: 6, 6, 6, 6*, 7, 9*, 10, 10*, 11*, 13, 16, 17*, 19*, 20*, 22, 23, 25*, 32*, 32*, 34*, 35*

Placebo 1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

The Kaplan-Meier estimators for these data are given in Tables 1 and 2 and its graphical representation may be seen in Figure 1.

TABLE 1. SURVIVAL FUNCTION FOR GROUP TREATED WITH 6-MP

t_i	6	7	10	13	16	22	23
$\hat{S}_x(t_i)$	0.86	0.81	0.75	0.69	0.63	0.54	0.45

TABLE 2. SURVIVAL FUNCTION FOR GROUP TREATED WITH PLACEBO

t_i	1	2	3	4	5	8	11
$\hat{S}_x(t_i)$	0.90	0.81	0.76	0.67	0.57	0.38	0.29
t_i	12	15	17	22	23		
$\hat{S}_x(t_i)$	0.19	0.14	0.09	0.05	0.00		

Kaplan-Meier estimates of survival
(Source: Freirich *et al.*, 1963)

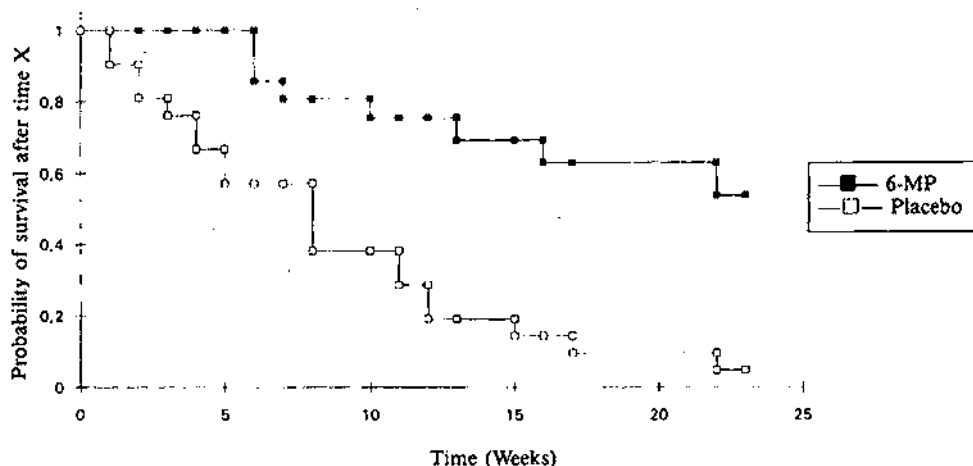


Figura 1. Estimated Survival Functions for Freirich's Data Set.

First applications of the bootstrap

a) Variance estimation

The Kaplan-Meier estimator of the survival function has very good properties: it is strongly consistent and it is asymptotically normal. This last fact

allows us to compute an estimation of the asymptotic variance at point t of this estimator. Greenwood's formula for the standard deviation is:

$$\hat{\sigma}_G(t) = \hat{S}_X^n(t) \sqrt{\sum_{U_i \leq t} \frac{\delta_i}{(n-i)(n-i+1)}}$$

However we can use the bootstrap to perform another estimation of this standard deviation. Efron (1981) suggests the following resampling plan: Given the sample:

$$(U_1, \delta_1), \dots, (U_n, \delta_n)$$

take a bootstrap sample (giving probability $1/n$ to each element of the previous sample):

$$(U_1^*, \delta_1^*), \dots, (U_n^*, \delta_n^*)$$

Efron shows that this plane is equivalent to resampling X_1^*, \dots, X_n^*

from X_1, \dots, X_n , and D_1^*, \dots, D_n^* from D_1, \dots, D_n and defining:

$$U_i^* = \min (X_i^*, D_i^*)$$

and

$$\delta_i^* = \begin{cases} 1, & \text{if } U_i^* = X_i^* \\ 0, & \text{if } U_i^* = D_i^* \end{cases}$$

Reid (1981) suggests another resampling plan, which consists of drawing samples following the Kaplan-Meier distribution, which assigns mass (not uniformly) only to uncensored observations.

Turning back to Efron's plan, we compute the Kaplan-Meier estimator for the bootstrap sample, $\hat{S}_X^*(t)$ and define the bootstrap estimator of the variance of $\hat{S}_X(t)$, which we may indicate $\hat{\sigma}_B$ as the standard deviation of $\hat{S}_X^*(t)$.

As usually (in a non-simple situation), this standard deviation may be approximated by the Monte Carlo method, taking B bootstrap samples, and computing the sample standard deviation of the corresponding Kaplan-Meier estimators at the point t . To be specific, let

$$\hat{S}_X^{*1}, \dots, \hat{S}_X^{*n}$$

be the corresponding estimations at t of the bootstrap samples. Then

$$\hat{\sigma}_B = \sqrt{\frac{\sum_{i=1}^B \hat{S}_X^{*i}(t)^2 - \frac{1}{B} \sum_{i=1}^B \hat{S}_X^{*i}(t)^2}{B-1}}$$

Greenwood's and Bootstrap estimate of the variance are shown in Table 3 for Freirrich's data set.

TABLE 3. GREENWOOD'S AND BOOTSTRAP (B=250) ESTIMATES OF THE VARIANCE OF $S(t)$

t_i	6	7	10	13	16	22	23
$V_G[\hat{S}_X(t_i)]$ (GreenWood)	.0058	.0076	.0093	.0114	.013	.0164	.0181
$V.[\hat{S}_X(t_i)]$ (Bootstrap)	.0056	.0069	.0083	.0107	.0126	.0154	.0178

b) Estimating the bias of the Median

The Kaplan-Meier estimator does not work well when t is near the highest values observed, and in particular, if the censoring is heavy. For this reason, and also because it is easily interpretable, the median is always an interesting parameter to estimate. On the other hand, there is no classical result (like the Greenwood formula) to estimate the bias and standard deviation of the estimation of the median (also computed by Kaplan-Meier). Once again, Efron (1981) suggests using the bootstrap.

The median is estimated by

$$\hat{m} = \inf \{t : \hat{S}_X^n(t) \leq 1/2\}$$

Now we take B bootstrap samples, compute the B medians, and the bias is estimated by

$$\hat{b} = \frac{1}{B} \sum_{i=1}^B \hat{m}_i^* - \hat{m}$$

and the standard deviation by the sample standard deviation as above.

EXAMPLE 2

The following data are part of a research conducted at the Mayo Clinic between 1974 and 1984, and the variable is the time (in days) passing between registration and liver transplantation, or moment of death or end of study. For the moment, we study only the patients taking a drug (D-penicillamine). We have to remark that the data are numerous, and there are many covariates which would allow fitting to a Cox model (see Fleming and Harrington, 1991).

The data are the following: (* denotes a censored observation)

19, 32, 57, 96, 143, 193, 194, 205, 275, 322, 341, 371, 383, 469, 512, 562*, 605*, 637*, 638*, 644*, 651*

The Kaplan-Meier estimator for this data gives:

TABLE 4. SURVIVAL FUNCTION FOR FLEMING AND HARRINGTON'S DATA SET

t_i	18	31	57	96	142	144
$\hat{S}_X(t_i)$.958	.917	.875	.833	.792	.750
t_i	192	194	204	275	322	328
$\hat{S}_X(t_i)$.708	.667	.625	.583	.542	.500
t_i	340	342	371	468	512	582
$\hat{S}_X(t_i)$.458	.417	.375	.333	.292	.243

The *Bootstrap estimate of the bias* of the median of the survival curve: is, following the previous formula:

$$320.14 - 328 = -7.86$$

Bias-Corrected estimate of the median
(Source: Fleming & Harrington 1991)

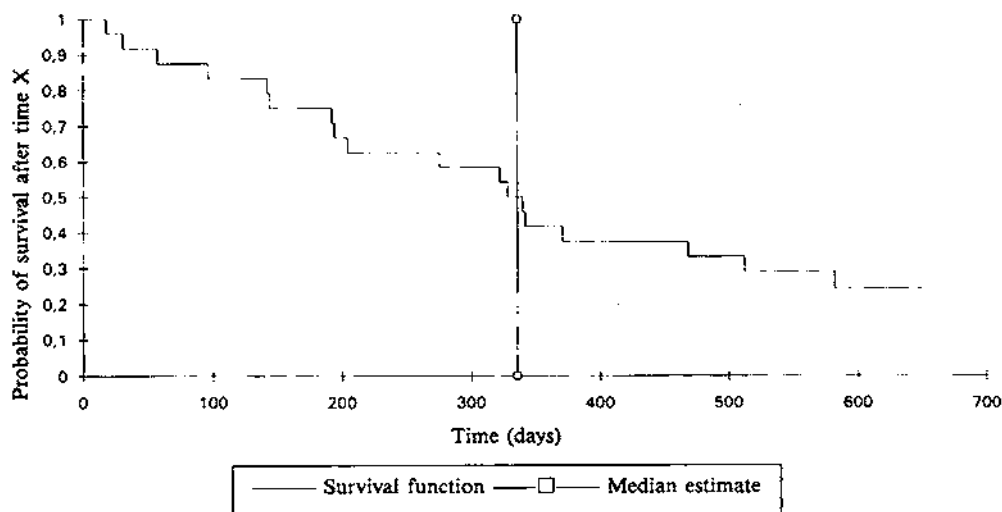


Figure 2. Bias Corrected Estimate of the Median.

Subtracting the bias from the sample median we may obtain a *bias corrected estimate of the median of the survival curve*

$$\text{Sample median} - \text{estimated bias} = 328 + 7.86 = 335.86$$

It is important to note that Efron gives a justification the use of bootstrap in these problems. However, a complete proof of the fact that the bootstrap works in this situation can be found in Gill (1989), where he shows that if the delta method works, the bootstrap works (at least in a weak sense).

Confidence bands for the survival function

a) Hall & Wellner's classical bands

Let γ be a confidence level. The problem of interest is now to find a region of the plane which contains the survival function with probability greater than or equal to γ . In the situation we are dealing with, this band is expressed in terms of the Kaplan-Meier estimator. We have to find a λ , such that

$$\hat{S}_x^n(t) - \lambda D_n(t) \leq S_x(t) \leq \hat{S}_x^n(t) + \lambda D_n(t), \forall t \in \mathbf{R}$$

with probability equal or greater than γ .

Assuming that the survival functions of X and D are continuous, Hall and Wellner (1980) constructed a confidence band using the asymptotic behaviour of the Kaplan-Meier estimator. Particularly they proved that

$$\lim_n \sup_{t \in [0, \tau]} \sqrt{n} \left| \hat{S}_x^n(t) - S_x(t) \right| \frac{1}{\hat{S}_x^n(t) (1 + C_n(t))} = \sup_{t \in [0, \tau]} |B^0(K(t))|$$

where τ is a point such that

$$S_x(\tau) > 0,$$

B^0 is a brownian bridge, and K is a function depending on the survival functions of X and D , and which is continuous with the hypotheses of this section. Then we have

$$\sup_{t \in [0, \tau]} |B^0(K(t))| = \sup_{t \in [0, K(\tau)]} |B^0(t)|$$

and we can construct conservative bands using the random variable

$$\sup_{t \in [0, 1]} |B^0(t)|$$

which is tabulated. To be specific, we choose λ such that

$$P(\sup_{t \in [0, 1]} |B^0(t)| \leq \lambda) = \gamma$$

In that case we have:

$$\hat{S}_X^n(t) - \lambda D_n(t) \leq S_X(t) \leq \hat{S}_X^n(t) - \lambda D_n(t), \forall t \in [0, \tau]$$

with probability greater than or equal to γ , where

$$D_n(t) = \sqrt{n} \hat{S}_X^n(t) (1 + C_n(t))$$

where

$$C_n(t) = n \sum_{U_i \geq t} \frac{\delta_i}{(n-i)(n-i+1)}$$

From this construction, we can deduce a test to determine whether the data are from a continuous survival function:

$$H_0 : S_X(t) = S^0(t), \forall t \in [0, \tau]$$

against a bilateral hypotheses

$$H_1 : \exists t \in [0, \tau] : S_X(t) \neq S^0(t)$$

using the test statistic:

$$J_n = \sup_{t \in [0, \tau]} \sqrt{n} \left| (\hat{S}_X^n(t) - S^0(t)) \frac{1}{\hat{S}_X^n(t) (1 + C_n(t))} \right|$$

and critical region

$$J_n > t_\alpha$$

where t_α is the $(1-\alpha)$ -percentile of the distribution of

$$\sup_{t \in [0, \tau]} |B^0(t)|$$

Asymptotically this test has level α ,

$$\lim_n P(J_n > t_\alpha | H_0) \leq \alpha$$

and it is consistent:

$$\lim_n P(J_n > t_\alpha | H_1) = 1$$

(This last fact is deduced from the strong consistency of the Kaplan-Meier estimator.)

b) Akrita's bootstrap bands

When no assumption on the continuity of the survivals of X and D is made, all the above steps are correct except

$$\sup_{t \in [0, \tau]} |B^0(K(t))| = \sup_{t \in [0, K(t)]} |B^0(t)|$$

In this general situation this means that the asymptotic distribution can depend strongly of the survival functions of X and D . The solution, then, is to bootstrap Hall & Wellner's confidence bands. This was carried out by Akritas (1986), who made use of the fact that, given γ , if we determine λ such that

$$P^* \left\{ \sqrt{n} \sup_{t \in [0, \tau]} |(\hat{S}_X^n(t) - \tilde{S}_X^n(t)) \frac{1}{\hat{S}_X^n(t) (1 + C_n(t))}| \leq \lambda \right\} = \gamma$$

where P^* denotes the bootstrap probability, then,

$$\hat{S}_X^n(t) - \lambda D_n(t) \leq S_X(t) \leq \hat{S}_X^n(t) + \lambda D_n(t), \forall t \in [0, \tau]$$

To determine the λ , we use the Monte Carlo method: We compute B times

Akrita's Bootstrap confidence bands
(Source: Freirrich *et al.*, 1963)

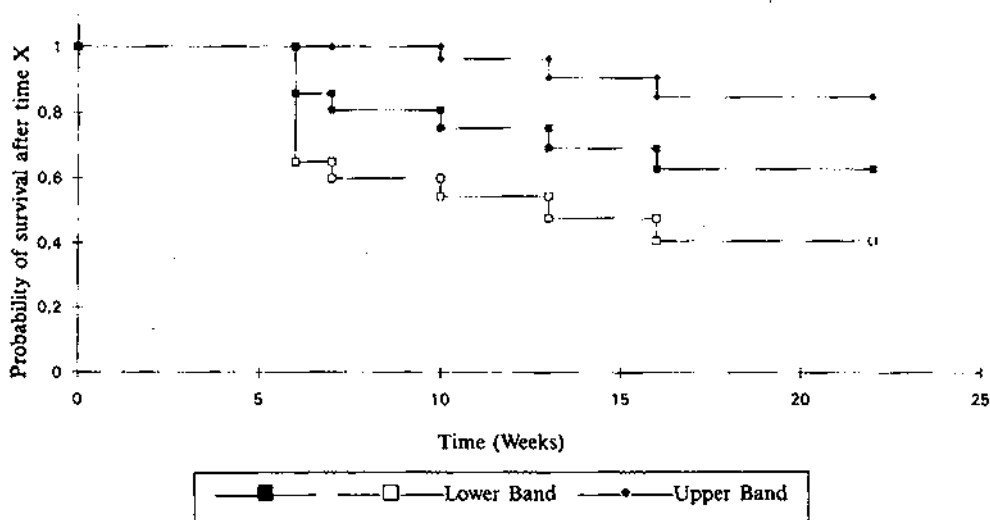


Figura 3. Bootstrap Confidence Bands for 6-MP.

$$\sqrt{n} \sup_{t \in [0, \tau]} |(\hat{S}_X^n(t) - \tilde{S}_X^n(t)) \frac{1}{\hat{S}_X^n(t) (1 + C_n(t))}|$$

and choose λ to be the γ percentile of those numbers.

TABLE 5. BOOTSTRAP CONFIDENCE BANDS (B=250) FOR SURVIVAL FUNCTION FOR GROUP TREATED WITH 6-MP

t_i	6	7	10	13	16	22	23
$\hat{S}_X(t_i)$.857	.807	.753	.690	0.627	0.538	0.45
Lower band	.649	.598	.542	.474	.406	.292	
Upper band	1.00	1.00	.963	.906	.848	.783	

Test to contrast the equality of survival functions

Consider two independent samples:

$$(X_1, D_1), \dots, (X_n, D_n)$$

and

$$(Y_1, G_1), \dots, (Y_m, G_m).$$

We would like to contrast

$$H_0 : S_X = S_Y$$

against an alternative, for example, the unilateral alternative

$$H_1 : S_X \geq S_Y$$

a) Classical test

Assume the survival functions of X , Y , D and G are continuous. We can construct the following Kolmogorov-Smirnov type test: Let

$$\hat{S}^{nm} \text{ and } C^{nm}$$

be the Kaplan-Meier estimator and the corresponding C of the pooled observations (that means, of all the observations). Consider the test statistic

$$J_{nm}^+ = \sqrt{\frac{nm}{n+m}} \sup_{t \in (0, \tau]} \frac{1}{\hat{S}^{nm}(1 + C^{nm}(t))} [\hat{S}_X^+(t) - \hat{S}_Y^+(t)]$$

then, assuming

$$\lim_{n,m \rightarrow \infty} \frac{m}{n+m} = r \in (0,1)$$

we have, under the null hypothesis,

$$\lim_{n,m \rightarrow \infty} J_{nm}^+ = \sup_{t \in [0,1]} B^0(K_X(t))$$

in distribution. Using the continuity conditions we have imposed, we can compute the distribution of the limit random variable. On the other hand, it is easy to show that under the alternative hypothesis,

$$\lim_{n,m \rightarrow \infty} J_{nm}' = +\infty$$

a.s., and it follows that we obtain a test which is asymptotically of level α , and is also consistent.

b) Bootstrap test

Akritis' results (1981), and the independence of the resamples, allows us to justify the use of a bootstrap test theoretically without assuming the continuity of the survival functions. To be specific,

given $0 < \alpha < 1$, let

$$\lambda_{n,m,\alpha}(\omega) = \inf \{ s : P^*(J_{nm}^{+*} \geq s) \geq \alpha \}$$

To test

$$H_0 : S_X = S_Y$$

against

$$H_1 : S_X \geq S_Y$$

consider the test with critical region

$$J_{nm}^+ > \lambda_{n,m,\alpha}$$

which has asymptotically level α , and is consistent. As before, to determine the critical point, we have to take B (n and m) bootstrap samples, compute B times

$$\sqrt{\frac{nm}{n+m}} \sup_{t \in [0,1]} \frac{1}{\hat{S}_{nm}(t) (1 + C^{nm}(t))} [(\hat{S}_X^{n*}(t) - \hat{S}_X^n(t)) - (\hat{S}_Y^{m*}(t) - \hat{S}_Y^m(t))]$$

and compute the $(1-\alpha)$ percentile of these numbers.

Once again if we compare the survival function for patients using Placebo and 6-MP we obtain the following results:

- Value of test statistic, $J_{nm} + = 1.9216$
- Critical value, $\lambda (B=250) \lambda = 0.3739$
- The decision criterion is: «Reject H_0 if $J_{nm} + > \lambda$ »

This results are clearly consistent with Figure 1 which shows that the 2 survival functions are quite different.

REFERENCES

- Akritis, M.G. (1986). Bootstrapping the Kaplan-Meier Estimator, *Journal of the American Statistical Association*, 81, 1032-1038.
- Efron, B. (1967). The Two Sample Problem with Censored Data, Proc. of the Fifth Berkeley Symp., IV, 831-853.
- Efron, B. (1981). Censored Data and the Bootstrap, *Journal of the American Statistical Association*, 76, 312-319.
- Fleming, T.R. and Harrington, D.P. (1991). Counting Processes and Survival Analysis, New York: Wiley.
- Gill, R.D. (1980). Censoring and Stochastic Integrals. Mathematish Centrum, Amsterdam.
- Gill, R.D. (1989). Non- and Semi-parametric Maximum Likelihood Estimates and the Von Mises Method (part I), *Scand. J. Stat.*, 16, 97-128.
- Hall, W.J. & Wellner, J.A. (1980). Confidence Bands for a survival curve from Censored Data. *Biometrika*, 67, 1, 133-143.
- Kaplan, E.L. & Meier, P. (1958). Non Parametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, 53, 457-481.
- Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data. New York: Wiley.
- Reid, N. (1981). Estimating the Median Survival Time. *Biometrika*, 67, 133-143.
- Romano, J.P. (1988). A Bootstrap Revival of some Nonparametric distance Test, *Journal of American Statistical Association*, 83, 698-708.
- Romano, J.P. (1989). Bootstrap and Randomization Test of some Non Parametric hypotheses. *Ann. Stat.*, 17, 141-149.
- Sánchez, A., Ocaña, J. & Ruiz de Villa, C. (1992). An Environment for Monte Carlo Simulation Studies (EMSS). 10th Symposium on Computational Statistics (COMPSTAT). Physica-Verlag. Springer Verlag.
- Schumacher, M. (1984). Two-sample Test of Cramer-Von-Mises and Kolmogorov-Smirnov Type for Randomly Censored Data. *Inst. Stat. Rev.*, 52, 262-281.

