Analysis of Censored Data
IMS Lecture Notes – Monograph Series (1995) Volume 27

THE STATISTICAL ANALYSIS OF KAPLAN-MEIER INTEGRALS *

Winfried Stute
University of Giesse

Abstract

Let $\hat{F}_n$ denote the Kaplan-Meier estimator computed from a sample of possibly censored data, and let φ be a given function. In this paper some of the most important properties of the Kaplan-Meier integral $\int \varphi d\hat{F}_n$ are reviewed.

1 Introduction

Statistical inference on the common mean of a set of independent identically distributed (i.i.d.) observations is dealt with in almost every textbook on statistical methodology. To name only a few facts, if $X_1, ..., X_n$ are i.i.d. random variables from some distribution function (d.f.) F, then the corresponding sample mean

$$S_n = n^{-1} \sum_{i=1}^{n} X_i$$

constitutes a consistent unbiased estimator of the unknown expectation $\mu := \int xF(dx)$ (assumed to exist):

$$\text{(1.1)} \qquad \mathbb{E}S_n = \mu \qquad \text{and} \qquad S_n \to \mu \quad \text{with probability one.}$$

The first statement is trivial while the second is just the SLLN. Moreover, under a finite second moment assumption, the CLT guarantees

$$\text{(1.2)} \qquad n^{1/2}[S_n - \mu] \to \mathcal{N}(0, \sigma^2) \qquad \text{in distribution,}$$

with

$$\sigma^2 = Var X_1 = \int x^2 F(dx) - \left[\int xF(dx)\right]^2.$$

*Supported by the "Deutsche Forschungsgemeinschaft".

1991 MSC. 62G05, 60F15, 62E20, 62G30, 62G15.
Key words and Phrases: Kaplan-Meier integral, SLLN, CLT, Bias, Variance, Jackknife.

(1.2) may be utilized, for example, to establish confidence intervals for μ. In more complicated setups, (1.1) and (1.2) are often needed for proper transformations of the X's rather than the X's themselves. For any such φ, we then get

$$(1.3) \quad S_n^\varphi \equiv n^{-1}\sum_{i=1}^n \varphi(X_i) \to \int \varphi dF \equiv S^\varphi \qquad \text{with probability one}$$

and

$$(1.4) \qquad n^{1/2}[S_n^\varphi - S^\varphi] \to \mathcal{N}(0,\sigma^2) \qquad \text{in distribution,}$$

where now

$$(1.5) \qquad \sigma^2 = \int \varphi^2 dF - \left[\int \varphi dF\right]^2.$$

For smooth but nonlinear statistical functionals $T(F_n)$ of the empirical d.f.

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \le x\}},$$

asymptotic normality of $T(F_n)$, for example, is obtained by expanding $T(F_n)$ into a linear part $\int \varphi dF_n = S_n^\varphi$ and a remainder, where now φ equals the influence function associated with T.

A typical feature which comes up in the analysis of lifetime data is censorship. Quite often, X represents the time elapsed from a patient's entry into a follow-up study until death. If at the time of statistical analysis the patient is still alive or withdrew from the study for some reason, the variable of interest will not be available. A convenient way to model this situation has been to introduce a random variable Y being independent of X such that only

$$(1.6) \qquad Z = \min(X,Y) \text{ and } \delta = 1_{\{X \le Y\}}$$

are observable. We shall refer to (1.6) as independent censorship. δ indicates whether X has been censored or not. Given a set $(Z_i,\delta_i), 1 \le i \le n$, of independent replicates of (Z,δ), it is then our goal to draw some inference on the true but unknown lifetime distribution F, while G, the d.f. of Y, is considered a nonparametric nuisance parameter.

Coming back to the case of completely observable data and recalling F_n, (1.4) becomes

$$n^{1/2}\int \varphi d(F_n - F) \to \mathcal{N}(0,\sigma^2).$$

Under random censoring it is tempting to estimate S^φ by

$$\hat{S}_n^\varphi = \int \varphi d\hat{F}_n,$$

in which $\hat{F}_n$ is a nonparametric substitute for F_n computable from the $(Z,\delta)'s$. Now, it is well known that the nonparametric maximum likelihood estimator of F is given by the Kaplan-Meier (1958) product-limit estimator defined by

$$1-\hat{F}_n(x)=\prod_{i=1}^{n}\left[1-\frac{\delta_{[i:n]}}{n-i+1}\right]^{1\{Z_{i:n}\le x\}}. \tag{1.7}$$

Here, $Z_{1:n}\le ... \le Z_{n:n}$ are the ordered Z-values, where ties within lifetimes or within censoring times are ordered arbitrarily and ties among lifetimes and censoring times are treated as if the former precedes the latter. $\delta_{[i:n]}$ denotes the concomitant associated with $Z_{i:n}$, i.e. $\delta_{[i:n]}=\delta_j$ if $Z_{i:n}=Z_j$. The Kaplan-Meier integral $\hat{S}_n^{\varphi}$ may then be written as

$$\hat{S}_n^{\varphi}=\sum_{i=1}^{n}W_{in}\varphi(Z_{i:n}),$$

where for $1\le i\le n$

$$W_{in}=\frac{\delta_{[i:n]}}{n-i+1}\prod_{j=1}^{i-1}\left[\frac{n-j}{n-j+1}\right]^{\delta_{[j:n]}}$$

is the mass attached to the i-th order statistic $Z_{i:n}$ under $\hat{F}_n$. When all δ's equal one, i.e., when all data are uncensored, each W_{in} becomes $\frac{1}{n}$ and therefore $\hat{S}_n^{\varphi}=S_n^{\varphi}$. Under censorship, however, $\hat{S}_n^{\varphi}$ is a complicated sum of functions of the Z-order statistics properly weighted by the random W's. Consequently results which are valid for sums of independent random variables are of no use for the analysis of $\hat{S}_n^{\varphi}$.

For further discussion note that (1.7) may also be motivated by incorporating a one-to-one relationship between a distribution function F and its pertaining cumulative hazard function

$$\Lambda_F(x)=\int_0^x\frac{F(dy)}{1-F(y-)}\equiv\int_{[0,x]}\frac{F(dy)}{1-F(y-)},$$

where in the following $F_-(y)\equiv F(y-)=\lim_{x\uparrow y}F(x)$ and $F\{y\}=F(y)-F(y-)$. Similarly for Λ_F. We then have, cf. Shorack and Wellner (1986), p. 301,

$$1-F(x)=e^{-\Lambda_c(x)}\prod_{\substack{a_i\in A\\ a_i\le x}}[1-\Lambda_F\{a_i\}], \tag{1.8}$$

with Λ_c denoting the continuous part of Λ_F and A being its atoms. Now, introduce H, the d.f. of Z, and set

$$\tilde{H}^1(x)=P(Z\le x,\delta=1).$$

By independence of X and Y,

$$(1-H) = (1-F)(1-G)$$

and

$$\tilde{H}^1(x) = \int_0^x (1-G(y-))F(dy).$$

Conclude that

$$\Lambda_F(x) = \int_0^x \frac{\tilde{H}^1(dy)}{[1-H(y-)]}. \tag{1.9}$$

H and $\tilde{H}^1$ may be consistently estimated by their empirical counterparts

$$H_n(y) = n^{-1} \sum_{i=1}^n 1_{\{Z_i \le y\}}$$

and

$$\tilde{H}_n^1(y) = n^{-1} \sum_{i=1}^n 1_{\{Z_i \le y, \delta_i = 1\}}.$$

Plugging them into the R.H.S. of (1.9) we obtain the Nelson-Aalen estimator of Λ_F:

$$\begin{aligned} \Lambda_n(x) &= \int_0^x \frac{\tilde{H}_n^1(dy)}{[1-H_n(y-)]} \\ &= \sum_{i=1}^n \frac{1_{\{Z_i \le x, \delta_i = 1\}}}{[n - Rank Z_i + 1]}. \end{aligned}$$

Since Λ_n is purely discrete, an application of (1.8) yields for the corresponding survival function

$$1 - \tilde{F}_n(x) = \prod [1 - \Lambda_n\{Z_i\}]^{1_{\{Z_i \le x, \delta_i = 1\}}}. \tag{1.10}$$

Check that $\tilde{F}_n = \hat{F}_n$.

It is the purpose of the present paper to review and discuss some of the most important properties of $\hat{S}_n^\varphi$, namely

Strong Consistency
Distributional Convergence
Bias
Jackknife

Finally, in the last section, we report on a small simulation study for $\varphi(x) = x$, i.e., the mean lifetime estimator.

Since our main emphasis will be on a general φ but most work on Kaplan-Meier integrals has been done for indicators, i.e., on the Kaplan-Meier estimator itself, this survey will be necessarily rather incomplete. This drawback may be excusable, however, since there already exist excellent monographs on the Kaplan-Meier estimator and its role in survival analysis; see Andersen et al. (1993), Gill (1994), Fleming and Harrington (1991) and Shorack and Wellner (1986). In contrast our choice of the material will mainly focus on

1. techniques which are designed to offer a powerful alternative to the usual counting process approach.

2. results which are valid without the model-assumption (1.6) of independent censorship.

The last point needs some further clarification. Though for the derivation of (1.10) the independence of X and Y was crucial, the Kaplan-Meier estimator (together with its integrals) makes sense also in the case when this assumption is violated. Actually,since we only observe the (Z, δ)'s, independent censorship can never be checked so that the investigation of $\hat{S}_n^\varphi$, irrespective of whether (1.6) is satisfied or not, becomes an important issue.

2 Strong Consistency

For indicators $1_{[0,x]}$, (1.8) applied to $\hat{F}_n$ and Λ_n allows for a reduction of the analysis of $\hat{F}_n(x)$ to that of the Nelson-Aalen estimator, which has a simpler structure. In particular, if one restricts oneself to $x \leq T < \tau_H$, where

$$\tau_H = inf\{x : H(x) = 1\} \leq \infty$$

is the least upper bound for the support of H, then the denominator in the integral defining Λ_n causes less troubles. Tools from classical empirical process theory may then be applied to obtain consistency with rates.

To be more precise, expand the integrand in $\Lambda_n(x)$ into

$$\frac{1}{1 - H_{n-}} = -\frac{1 - H_{n-}}{(1 - H_-)^2} + \frac{2}{1 - H_-} + \frac{(H_{n-} - H_-)^2}{(1 - H_-)^2(1 - H_{n-})}.$$

By the LIL for empirical d.f.'s, the last term tends to zero as $n^{-1} \ln \ln n$ with probability one and uniformly in $x \leq T < \tau_H$. Consequently

$$(2.1) \qquad \Lambda_n(x) = -\int_0^x \frac{1 - H_n(y-)}{(1 - H(y-))^2} \tilde{H}_n^1(dy) + 2 \int_0^x \frac{\tilde{H}_n^1(dy)}{1 - H(y-)} + O(n^{-1} \ln \ln n).$$

By the SLLN and a standard uniformity argument, we obtain for the second integral with probability one

$$\lim_{n\to\infty} \sup_{x\leq T} \left| \int_0^x \frac{\tilde{H}_n^1(dy)}{1-H(y-)} - \int_0^x \frac{\tilde{H}^1(dy)}{1-H(y-)} \right| = 0. \tag{2.2}$$

As to the first integral in (2.1), a combination of Glivenko-Cantelli (for H_n) together with (2.2) leads to

$$\lim_{n\to\infty} \sup_{x\leq T} \left| \int_0^x \frac{1-H_n(y-)}{(1-H(y-))^2} \tilde{H}_n^1(dy) - \int_0^x \frac{\tilde{H}^1(dy)}{1-H(y-)} \right| = 0$$

so that in summary

$$\lim_{n\to\infty} \sup_{x\leq T} |\Lambda_n(x) - \Lambda(x)| = 0 \qquad \text{w.p.1.} \tag{2.3}$$

Replacing Glivenko-Cantelli by a proper LIL (2.3) may be improved to

$$\sup_{x\leq T} |\Lambda_n(x) - \Lambda(x)| = O\left(\sqrt{\frac{\ln\ln n}{n}}\right) \qquad \text{w.p.1.}$$

We have discussed the derivation of (2.3) in somewhat greater detail not because it is very exciting. It has been included mainly to show that in view of existing results for empirical d.f.'s the uniform convergence on compacta of the Nelson-Aalen estimator is obtained almost for free.

On the other hand, since Λ_n is bounded but Λ is unbounded whenever F is continuous, we cannot expect uniform convergence on the whole real line in this case. The best we can hope for in general is that $\Lambda_n \to \Lambda$ (with or without rates) uniformly on $x \leq T_n$, where $T_n \to \tau_H$ as $n \to \infty$ at appropriate rates. Needless to say that since now $1-H_n$ is no longer bounded away from zero, the above reasoning requires some serious modifications. See Stute (1994a).

Recall that in this paper, rather than Λ_n, our main interest is in the Kaplan-Meier estimator (integral). The last somewhat pessimistic remarks about the uniform convergence of Λ_n to Λ may lead one to believe that the same is true for $\hat{F}_n$ and F. Actually, the problems with the right tails occur only if one traces the analysis of $\hat{F}_n - F$, via (1.8) and (1.10), back to that of $\Lambda_n - \Lambda$. In Stute and Wang (1993) a new approach has been proposed which does not utilize the cumulative hazard function as a vehicle to study $\hat{F}_n$. This new approach has the advantage that

(i) φ may be a general F-integrable function rather than an indicator

(ii) the crucial assumption (1.6) of independent censorship becomes superfluous in the sense that convergence of $\hat{S}_n^\varphi$ can be shown to hold (with probability one and in the mean) without (1.6). Moreover the limit equals the target value in case (1.6) is true.

The idea is as follows: instead of utilizing Λ_n it is tempting to look at

$$\hat{S}_n^\varphi = \sum_{i=1}^n W_{in}\varphi(Z_{i:n})$$

itself. At this stage it is worthwhile recalling the different techniques which are available to prove the SLLN for ordinary sample means:

(a) Kolmogorov's original proof and Etemadi's (1981) beautiful modification. Both proofs heavily rely on the fact that $S_n^\varphi = n^{-1}\sum\varphi(X_i)$ is a normalized sum of i.i.d. random variables.

(b) Recalling that an i.i.d. sequence is strictly stationary and ergodic the SLLN is also implied by the ergodic theorem.

(c) Note that for a proper sequence of σ-fields $(S_n^\varphi)_n$ is a reverse time martingale. So the martingale convergence theorem applies. That the limit is constant follows from the 0-1 law.

As may be expected the arguments needed for (a) and (b) cannot be extended to handle $\hat{S}_n^\varphi$. Also, as to (c), it can be seen that under censorship $\mathbb{E}\hat{S}_n^\varphi$ may differ from n to n. Consequently there is no hope that in general $\hat{S}_n^\varphi$ is a reverse martingale in n. Though this looks pessimistic it turns out, fortunately enough, that $\hat{S}_n^\varphi$ still carries a rich structure so as to make standard martingale theory applicable. For this, let

$$\mathcal{F}_n = \sigma(Z_{i:n}, \delta_{[i:n]}, 1 \le i \le n, Z_{n+1}, \delta_{n+1}, \ldots).$$

Then, clearly, $\hat{S}_n^\varphi$ is adapted to $\mathcal{F}_n$ with $\mathcal{F}_n \downarrow \mathcal{F}_\infty$, say. Moreover, by the Hewitt-Savage zero-one law, $\mathcal{F}_\infty$ is trivial. The following equation is taken from Lemma 2.2 in Stute and Wang (1993):

$$\mathbb{E}(\hat{S}_n^\varphi|\mathcal{F}_{n+1}) = \hat{S}_{n+1}^\varphi - R_{n+1}, \tag{2.4}$$

where

$$R_{n+1} = \frac{1}{n+1}\varphi(Z_{n+1:n+1})\delta_{[n+1:n+1]}(1-\delta_{[n:n+1]})\prod_{j=1}^{n-1}\left[\frac{n-j}{n-j+1}\right]^{\delta_{[j:n+1]}}$$

As a first trivial consequence

$$\mathbb{E}(\hat{S}_n^\varphi|\mathcal{F}_{n+1}) \le \hat{S}_{n+1}^\varphi \qquad \text{for } \varphi \ge 0,$$

i.e. $(\hat{S}_n^\varphi)_n$ is a reverse-time supermartingale. Proposition 5-3-11 in Neveu (1975) immediately yields convergence with probability one and in the mean of $(\hat{S}_n^\varphi)_n$ when $\varphi \geq 0$. Decompose φ into its positive and negative part when it attains both signs to handle a general φ. By Hewitt-Savage the limit must be a constant. To state the result note that if we don't assume independent censorship (1.6) there will be no Y and hence no G. It is thus necessary to formulate the result in terms of quantities which uniquely determine the distribution of the observed (Z, δ)'s. We already introduced H, the d.f. of Z. If we put

$$m(x) = \mathbb{P}(\delta = 1 | Z = x),$$

the joint distribution of (Z, δ) is completely determined by H and m. Putting

$$\gamma_0(x) = \int_0^{x-} \frac{1 - m(y)}{1 - H(y)} H(dy).$$

Lemma 2.7 of Stute and Wang (1993) (formulated there only for a continuous H) asserts that the limit of $\hat{S}_n^\varphi$ equals

$$S = \int \varphi(x) m(x) \gamma_0(x) H(dx), \tag{2.5}$$

so that, under $\int |\varphi| dF < \infty$,

$$\lim_{n \to \infty} \hat{S}_n^\varphi = S \qquad \text{with probability one and in the mean.} \tag{2.6}$$

Under independent censorship S becomes

$$S = \int_{\{x < \tau_H\}} \varphi(x) F(dx) + 1_{\{\tau_H \in A\}} \varphi(\tau_H) F\{\tau_H\}, \tag{2.7}$$

where A is the set of H-atoms (possibly empty).

(2.6) is the extension of the SLLN to the case of general censorship. The discussion in Stute and Wang (1993) also shows that in many situations S in (2.7) equals the target value S^φ. Originally, the result had been proved under the additional assumption that F and G have no jumps in common, which is enough for practical purposes. An extension to the general case is possible, however, by incorporating a new time scale, similar to Stute (1995) in derving the CLT for Kaplan-Meier integrals. As pointed out previously in practice it is not possible to check the validity of (1.6). (2.5) and (2.6) may then be useful in a simulation study to find out in selected situations how much $\hat{S}_n^\varphi$, S and S^φ may differ when (1.6) is violated.

We only mention that the SLLN for $\hat{S}_n^\varphi$ is the key tool to prove consistency also for more involved estimators, like M-, L- or minimum distance estimators, under random censorship.

3 Distributional Convergence

In their landmark paper Kaplan and Meier (1958) not only derived the formula for the product-limit estimator $\hat{F}_n(x)$, but also added - on heuristic grounds - some useful comments on the (limit) covariance of $\hat{F}_n(x)$ and $\hat{F}_n(y)$. In particular, they pointed out that the variance has many similarities with Greenwood's (1926) and Irwin's (1949) formula in connection with actuarial estimates. They also mention (p. 476) that "in the derivation of approximate formulas any bias that $\hat{F}_n(x)$ may have is neglected". For the limit covariance of $\hat{F}_n(x)$ and $\hat{F}_n(y)$ (properly standardized) they come up with the expression (in our terms)

$$(3.1) \qquad (1-F(x))(1-F(y))\int_0^x \frac{1-G}{(1-H)^2}dF, \qquad x \leq y.$$

Since upon integrating by parts

$$\mu = \mathbb{E}X = \int_0^\infty [1-F(x)]dx$$

they proposed to estimate μ by

$$(3.2) \qquad \mu_n = \int_0^\infty [1-\hat{F}_n(x)]dx,$$

at least in cases when the last observation is uncensored. In this situation $\hat{F}_n$ is a proper d.f. so that μ_n is well defined. Writing μ_n^2 as a double integral and then using (3.1) they argued that the limit variance of μ_n equals

$$(3.3) \qquad \int_0^\infty \left[\int_x^\infty (1-F(v))dv\right]^2 \frac{F(dx)}{(1-H(x))(1-F(x))}.$$

In the context of distributional convergence (3.1) was first justified by Breslow and Crowley (1974). In their extension of Donsker's invariance principle for the empirical process, they showed that the Kaplan-Meier process

$$\hat{\alpha}_n(x) = n^{1/2}[\hat{F}_n(x) - F(x)], \qquad 0 \leq x \leq T < \tau_H$$

weakly converges in the Skorokhod space $D[0,T]$ to a centered Gaussian process with covariance as given in (3.1). The technique elaborated in section 7 of their paper became by now standard and was adopted by many authors in subsequent work. A somewhat different approach was presented in Burke, Csörgő and Horváth (1981, 1988). They derived a strong approximation of $\hat{\alpha}_n, n \geq 1$, by a sequence of Gaussian processes, in the spirit of Komlós, Major and Tusnády (1975). Lo and Singh (1986) and Major and Rejtö (1988) obtained almost sure representations of $\hat{\alpha}_n$ in terms of sums of independent processes (plus a remainder). See also Stute (1994a). This method requires a deeper analysis of the first integral appearing in (2.1), being a U-statistic process of degree two rather than a (simple) sum of independent quantities.

From our discussion in the previous section it becomes apparent why restriction to compact intervals $[0, T]$ with $T < \tau_H$ was essential. Gill (1983) was the first to establish weak convergence on the whole real line, under some natural technical assumptions guaranteeing that censoring effects do not dominate the variable of interest in the extreme right tails. See also Ying (1989). Their method of proof was based on by now well-known martingale techniques elaborated in the context of survival analysis in Gill (1980).

As to general Kaplan-Meier integrals much less has been known for a long time. Sander (1975), in discussing (3.2), came to the conclusion that "it is extremely difficult to obtain the distribution theory for the estimators of $\int_0^T (1 - F(x))dx$ whenever $T = \infty$". Susarla and Van Ryzin (1980) apparently were the first to provide a rigorous treatment for the mean lifetime estimator truncated at M_n, but such that $M_n \to \infty$ at appropriate rates, as $n \to \infty$. Gill (1983) applied convergence of the Kaplan-Meier process plus integration by parts to obtain, under some additional tail assumptions on the censoring mechanism, distributional convergence of the mean lifetime estimator and Kaplan-Meier integrals for φ's which are nonnegative, continuous and nonincreasing. Schick, Susarla and Koul (1988) obtained, for this class of φ's, a weak representation of $\int \varphi d\hat{F}_n$ in terms of a sum of i.i.d. random variables plus a remainder. In all of these papers integration by parts was essential. Yang (1994) was able to extend distributional convergence of $\int \varphi d\hat{F}_n$, under regularity conditions on F, to those φ's satisfying

$$\int \frac{\varphi^2}{1 - G} dF < \infty. \tag{3.4}$$

The integral (3.4) becomes part of the limit variance so that (3.4) is indispensable. Stute (1995) obtained a representation of $\int \varphi d\hat{F}_n$ as a sum of i.i.d. random variables plus a remainder which is valid under no regularity assumptions on F and G. Moreover, the paper was written within the framework (1.6) only as a historical tribute but may be readily extended to the case of general censorship, in which as in section 2 the distributional characteristics of the observed (Z, δ)'s are H and m and not F and G. The key observation of our approach is the fact that the Kaplan-Meier integral

$\int \varphi d\hat{F}_n$ may be written as

$$\sum_{i=1}^{n} W_{in}\varphi(Z_{i:n})$$
$$= \int \varphi(w) \exp\left\{ n \int_0^{w-} \ln\left[1 + \frac{1}{n(1-H_n(z))}\right] \tilde{H}_n^0(dz) \right\} \tilde{H}_n^1(dw)$$

in which $\tilde{H}_n^1$ is defined as in the previous section and correspondingly,

$$\tilde{H}_n^0(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{Z_i \leq x, \delta_i = 0\}}.$$

Expansion of the logarithmic term and neglecting error terms leads one to a U-statistic of degree three. Its Hájek projection is the desired (simple) sum of i.i.d. random variables to which the ordinary CLT applies. For a large class of φ's the error terms are $o(1)$ with probability one so that an application of the LIL also establishes the law of the iterated logarithm for Kaplan-Meier integrals.

The general formula for the limit variance of $\int \varphi d\hat{F}_n$ was discussed in Stute (1994d). For the purpose of the present paper it is enough to consider independent censorship (1.6) with a continuous F. We then have

$$n^{1/2}\left[\int \varphi d\hat{F}_n - \int_0^{\tau_H} \varphi dF\right] \longrightarrow \mathcal{N}(0, \sigma_1^2) \qquad \text{in distribution} \tag{3.5}$$

with

$$\sigma_1^2 = \int_0^{\tau_H} \frac{\varphi^2}{1-G} dF - \left[\int_0^{\tau_H} \varphi dF\right]^2 - \int \left[\int_x^{\tau_H} \varphi dF\right]^2 \frac{1-F(x)}{[1-H(x)]^2} G(dx). \tag{3.6}$$

Note that in this three-terms formula for σ_1^2, the last term vanishes if there is no censorship ($G \equiv 0$) so that in this case σ_1^2 reduces to σ^2 from (1.5):

$$\sigma_1^2 = \sigma^2 = \int \varphi^2 dF - \left[\int \varphi dF\right]^2.$$

4 Bias

We heard in section 2 that the bias of $\hat{F}_n(x)$ was already briefly discussed in Kaplan and Meier (1958). A first rigorous treatment may be found in Gill's (1980) thesis. In his formula (3.2.17) he showed that

$$-F(x)H^n(x) \leq \mathbb{E}\hat{F}_n(x) - F(x) \equiv Bias\hat{F}_n(x) \leq 0$$

i.e., $\hat{F}_n(x)$ is always biased downwards. What the left inequality also suggests is that the bias increases as x gets large and that it may then become a nonnegligible quantity. Mauro (1985) extended the right inequality to a general Kaplan-Meier integral:

$$Bias\left[\int \varphi d\hat{F}_n\right] \leq 0 \qquad \text{for } \varphi \geq 0.$$

Zhou (1988) was able to also establish a lower bound whenever $\varphi \geq 0$ is continuous and Riemann integrable:

$$-\int \varphi(t)H^n(t)F(dt) \leq Bias\left[\int \varphi d\hat{F}_n\right].$$

In Stute (1994c) we were able to derive a formula and an expansion for the bias in such a way that (see p. 476):

(a) unbiasedness was readily recovered if there is no censorship

(b) Bias $\to 0$ as $n \to \infty$ for a general φ

(c) sharp lower and upper bounds are easily available

(d) all expressions only depend on the joint distribution of the observed (Z, δ)'s

(e) the effect of light, medium of heavy censoring on the bias may be easily discovered.

The solution to this program is surprisingly complicated. But what comes out is that informally speaking the bias of a Kaplan-Meier integral may

- be zero	,	if there is no censoring
- decrease to zero exponentially fast	,	if, e.g., φ is bounded and vanishes right of some $T < \tau_H$.
- decrease to zero at any polynomial rate	,	if, e.g., $0 \leq \varphi(x) \uparrow \infty$ as $x \to \infty$ and censoring is heavy.

In particular, the bias may decrease to zero at a rate slower than $n^{-1/2}$ and therefore becomes an important quantity in assessing the quality of the approximation of $\int \varphi dF$ by $\int \varphi d\hat{F}_n$.
For indicator functions $\varphi = 1_{[0,t]}$, the reduction of the bias has been the subject of some discussion before. See Chen et al. (1982) and Wellner (1985). Chen et al. (1982) proposed to replace $\hat{F}_n$ by

$$F_n^*(y) = \begin{cases} \hat{F}_n(y) & \text{for} \quad y < Z_{n:n} \\ 1 & \text{for} \quad y \geq Z_{n:n} \end{cases}$$

Note that F_n^* reduces to $\hat{F}_n$ if we artificially set $\delta_{[n:n]} = 1$, irrespective of whether $Z_{n:n}$ has been censored or not. Wellner (1985) compared $\hat{F}_n$ with F_n^* and was led to prefer $\hat{F}_n$, because in the cases investigated by him the upward bias of F_n^* was worse than the downward bias of $\hat{F}_n$. Stute (1994b) proposed another modification of $\hat{F}_n$ which is based on the following observation. Suppose that all $X's$ were observable. Then the empirical d.f. based estimate of Λ_F would be

$$\Lambda_n(x) = \sum_{i=1}^{n} \frac{1_{\{X_{i:n} \leq x\}}}{n-i+1}. \tag{4.1}$$

In order to measure the impact of censoring under(1.6) we compute the conditional expectation of the Nelson-Aalen estimator w.r.t. the ordered $X's$. It then turns out that

$$\mathbb{E}\left[\Lambda_n(x) | X_{i:n}, 1 \leq i \leq n\right] = \sum_{i=1}^{n} \frac{1_{\{X_{i:n} \leq x\}}}{n-i+1} [1 - G^{n-i+1}(X_{i:n})].$$

Compared with (4.1) we see that censoring causes an additional bias term

$$\sum_{i=1}^{n} \frac{1_{\{X_{i:n} \leq x\}}}{n-i+1} G^{n-i+1}(X_{i:n}). \tag{4.2}$$

This is particularly large if G compared with F has short tails. Since (4.2) is unknown one may be tempted to replace G by its Kaplan-Meier estimator $\hat{G}_n$ and then to substitute the (unknown) X-sample by some bootstrap replicates $X_1^*, ..., X_n^*$ from F_n^*. Utilizing (1.8) again we finally come up with the following modified version of $\hat{F}_n$:

$$\begin{aligned} & 1 - \hat{F}_n^1(x) \\ = & \prod_{i=1}^{n} \left[1 - \frac{1_{\{Z_{i:n} \leq x, \delta_{[i:n]} = 1\}}}{n-i+1} \right] \\ \times & \prod_{i=1}^{n} \left[1 - \frac{1_{\{X_{i:n}^* \leq x\}} \hat{G}_n^{n-i+1}(X_{i:n}^*)}{n-i+1} \right]. \end{aligned}$$

Since $\hat{F}_n$ and F_n^* only jump at the Z's, we obtain for some weights W_{in}^1:

$$\hat{F}_n^1(x) = \sum_{i=1}^n W_{in}^1 1_{\{Z_{i:n} \leq x\}}.$$

Likewise,

$$\hat{S}_n^{1\varphi} \equiv \int \varphi d\hat{F}_n^1 = \sum_{i=1}^n W_{in}^1 \varphi(Z_{i:n}).$$

It is easily seen that $W_{in}^1 \geq W_{in}$ so that for $\varphi \geq 0$ the modified procedure reduces the downward bias of $\hat{S}_n^\varphi$. The difference between W_{in}^1 and W_{in} becomes negligible for small to moderate i while for $i = n, n-1, \ldots$ there may be a difference resulting in an upweighing of the extreme order statistics. The asymptotic theories for $\hat{S}_n^\varphi$ and $\hat{S}_n^{1\varphi}$ are the same. For finite sample size, Stute (1994b) pointed out through an extensive simulation study that $\hat{S}_n^{1\varphi}$ may have a significantly smaller bias and, somewhat unexpectedly,also a smaller variance.
In section 6 the confidence intervals for the mean lifetime were centered at $\hat{S}_n^{1\varphi}$ rather than $\hat{S}_n^\varphi$.

8 The Jackknife

The jackknife has been proposed to serve two purposes, see Quenouille (1956) and Tukey (1958):

(i) If T_n happens to be a biased statistic, the jackknife is expected to provide a modification of T_n with a smaller bias. For later reference, let $T_n = S(F_n)$ be a statistical functional evaluated at the (ordinary) empirical distribution function F_n. Denote with $F_n^{(k)}$ the empirical d.f. of the sample $X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n$ and put

$$\bar{T}_n := n^{-1} \sum_{k=1}^n S(F_n^{(k)}).$$

Then the bias-corrected substitute for T_n is defined as

$$\tilde{T}_n = T_n - (n-1)\{\bar{T}_n - T_n\}.$$

(ii) In the above notation the jackknife estimate of variance of T_n is defined as

$$\widehat{Var}(Jack) = \frac{n-1}{n} \sum_{k=1}^n [T_n^{(k)} - \bar{T}_n]^2$$

with

$$T_n^{(k)} = S(F_n^{(k)}).$$

A general account of the jackknife may be found in Gray and Schucany (1972) and Efron and Tibshirani (1993). For $T_n = \int \varphi dF_n$, there is no need for a bias correction. This is also confirmed by the jackknife, in view of $\tilde{T}_n = T_n$. The jackknife estimate of variance equals n^{-1} times the sample variance, which is one would expect.

Note that the crucial thing about (i) and (ii) is that the statistic of interest is a function of F_n and therefore attaches mass $1/n$ to each of the data. As a consequence deletion of one point just results in a change of the mass $1/n$ to $1/(n-1)$. For the Kaplan-Meier integral the situation is completely different since now the statistic is a sum of (functions of) order statistics weighted by the complicated random W_{in}'s. Denoting with $\hat{F}_n^{(k)}$ the Kaplan-Meier estimator from the entire sample except $(Z_{k:n}, \delta_{[k:n]})$, then $S(\hat{F}_n^{(k)})$ not only involves changes of the standard weights, but also incorporates replacement of the weights W_{in} by new ones depending on the labels $\delta_{[i:n]}, 1 \leq i \leq n$. This may be one of the reasons why the jackknife under random censorship has been dealt with only in few papers. Gaver and Miller (1983) proved that the jackknife corrected Kaplan-Meier estimator at a fixed $x < \tau_H$ has the same limit distribution as $\hat{F}_n(x)$. Stute and Wang (1994) derived, for an arbitrary φ, a finite sample formula for the jackknife modification of $\int \varphi d\hat{F}_n$:

$$(5.1)\quad \tilde{S}_n^{\varphi} = \hat{S}_n^{\varphi} + \frac{n-1}{n}\varphi(Z_{n:n})\delta_{[n:n]}(1-\delta_{[n-1:n]})\prod_{j=1}^{n-2}\left[\frac{n-1-j}{n-j}\right]^{\delta_{[j:n]}}.$$

Hence the correction term depends on the largest Z-observation only but on all δ-concomitants. Also the jackknife is much more cautious about attaching masses to the last observation when it is censored than what has been recommended in the ad-hoc proposal leading to F_n^* in the previous section. It is also worthwhile to compare the correction term in (5.1) with R_{n+1} in (2.4). First, both vanish unless the largest observation is uncensored and the second last is censored. Only in this case the extreme right data contain enough information on F to make a slight change of W_{nn} desirable.
As to the variance, (3.5) suggests that

$$Var[\hat{S}_n^{\varphi}] \sim \frac{\sigma_1^2}{n} \text{ as } n \to \infty.$$

So what $\widehat{Var}(Jack)$ is expected to do is

$$(5.2)\qquad n\widehat{Var}(Jack) \to \sigma_1^2 \qquad \text{with probability one.}$$

Since (omitting φ)

$$\text{(5.3)} \qquad n\widehat{Var}(Jack) = (n-1)\sum_{k=1}^{n}[S_n^{(k)}]^2 - n(n-1)\bar{S}_n^2.$$

and

$$\begin{aligned} S_n^{(k)} &= \sum_{i=1}^{k-1} \frac{\varphi(Z_{i:n})\delta_{[i:n]}}{n-i} \prod_{j=1}^{i-1}\left[\frac{n-j-1}{n-j}\right]^{\delta_{[j:n]}} \\ &+ \sum_{i=k+1}^{n} \frac{\varphi(Z_{i:n})\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{k-1}\left[\frac{n-j-1}{n-j}\right]^{\delta_{[j:n]}} \prod_{j=k+1}^{i-1}\left[\frac{n-j}{n-j+1}\right]^{\delta_{[j:n]}} \end{aligned}$$

already is a complicated expression to be squared in (5.3) it is a priori not obvious at all if $\widehat{Var}(Jack)$ is able to work out the three terms in the expression (3.6) for σ_1^2. In Stute (1994d) a finite sample formula for $\sum_{k=1}^{n}[S_n^{(k)}]^2$ was derived from which one obtains that up to a complicated error term

$$\begin{aligned} n\widehat{Var}(Jack) &= (n-1)\sum_{i=1}^{n-1} \varphi^2(Z_{i:n})\delta_{[i:n]}\frac{1}{(n-i)^2}\prod_{j=1}^{i-1}\left[\frac{n-j-1}{n-j}\right]^{2\delta_{[j:n]}} \\ &- \frac{n}{n-1}\hat{S}_n^2 \\ &- (n-1)\sum_{j=1}^{n-2}(1-\delta_{[j:n]})b_j \prod_{k=1}^{j-1}\left[\frac{(n-k-1)(n-k+1)}{(n-k)(n-k)}\right]^{2\delta_{[k:n]}} \\ &\qquad \times \left[\sum_{i=j+1}^{n}\varphi(Z_{i:n})W_{in}\right]^2, \end{aligned}$$

where

$$b_j = b_{jn} = \frac{1}{(n-j-1)^2} - \frac{1}{(n-j)^2} + \frac{1}{(n-j-1)} - \frac{1}{n-j}.$$

This somewhat mysterious representation of $n\widehat{Var}(Jack)$ in fact constitutes the empirical analog of the three-terms expression (3.6) for σ_1^2. By the SLLN for Kaplan-Meier integrals the second term converges to $-S^2$. The first and third expressions need some special care. But after all it can in fact be shown that they converge to the desired limits. Provided the error term is negligible these pieces altogether would imply (5.2). Unfortunately, and somewhat unexpectedly, this holds true only when $\varphi(x) \to 0$ as $x \to \tau_H$. So, in particular, the Jackknife yields a consistent estimate of σ_1^2 for $\varphi = 1_{[0,t]}$ with $t < \tau_H$. For a general φ, it may be inconsistent due to the observation that the remainder term becomes nonnegligible iff $|\varphi(Z_{n:n})|$ is moderately large and

$$\text{(5.4)} \qquad \delta_{[n-1:n]} = 0 \text{ and } \delta_{[n:n]} = 1.$$

If, under (5.4), we redefine $\delta_{[n:n]}$ by artificially putting $\delta_{[n:n]} = 0$, we obtain a slightly changed version of $\widehat{Var}(Jack)$, for which the remainder term vanishes and the three leading terms still converge. In other words, this modification satisfies (5.2) under the only assumption that σ_1^2 is finite.

6 Data Analytic Aspects Of Kaplan-Meier Integrals

In this section we will consider estimation of the mean lifetime, i.e. $\varphi(x) = x$. This is an important simple example of a φ- function which does not vanish right of some $T < \tau_H$. Hence all of the data are needed to compute the Kaplan-Meier integral and not just those which are bounded away from (the unknown) τ_H. Moreover, since φ is nondecreasing large values of Z will give a significant contribution to the value of the estimator.
In our simulation study only exponentially distributed variables were considered, namely

$$(6.1) \qquad 1 - F(x) = exp(-x) \qquad \text{and} \qquad 1 - G(x) = exp(-\lambda x)$$

for $x \geq 0$, with varying λ's. Clearly,

$$S^{\varphi} = \int xF(dx) = 1 \qquad \text{and} \qquad \int xG(dx) = 1/\lambda.$$

Under (6.1), Z and δ are independent with

$$\mathbb{P}(\delta = 1) = \frac{1}{1+\lambda}.$$

The following choices of λ were investigated in detail:

λ	$\lambda = 1/5$	$\lambda = 1/3$	$\lambda = 1/2$
Percentage of censored data	16,66 %	25 %	33,33 %

The nominal confidence level was $1 - \alpha = 0.95$. Sample sizes were $n = 10, 30, 50$ and 100. As mentioned in section 4, confidence intervals were centered at $\hat{S}_n^{1\varphi}$. For each sampling situation the number of runs was 200.

The table below lists the actual percentages of times the true parameter 1 was contained in

$$\tilde{I} := \left[\hat{S}_n^{1\varphi} - n^{-1/2}\hat{\sigma}_n\mu_{1-\alpha/2}, \hat{S}_n^{1\varphi} + n^{-1/2}\hat{\sigma}_n\mu_{1-\alpha/2}\right].$$

Here $\hat{\sigma}_n^2$ is the plug-in estimator of the limit variance σ_1^2 in (3.6) and $\mu_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution.

	$n = 10$	$n = 30$	$n = 50$	$n = 100$
$\lambda = 1/5$	0.87	0.92	0.93	0.94
$\lambda = 1/3$	0.79	0.85	0.93	0.94
$\lambda = 1/2$	0.73	0.86	0.93	0.95

It becomes apparent that

- for a given n the actual coverage percentages decrease as censoring effects become substantial
- for $n \geq 50$ they are almost constant and very close to the nominal level
- for $n \leq 30$ the results are satisfactory for $\lambda = 1/5$
- for $n = 10$ and $\lambda = 1/3$ and $\lambda = 1/2$ the loss of information due to censoring results in an unstable estimate with a less satisfactory behavior.

After all one can say that even for small sample sizes the proposed confidence intervals enjoy excellent coverage properties if censoring is not too heavy ($\lambda = 1/5$). For $\lambda \geq 1/3$, coverage becomes excellent already for $n \geq 50$.

It is also very instructive to look at the confidence intervals themselves rather than only reporting on the coverage percentages. Below, for some selected n and λ, the values of 100 $\hat{S}_n^{1\varphi}$'s together with the pertaining confidence intervals are presented. For the sake of illustration, also the case of "no censorship" is considered. For $n = 50$ and $\lambda = 1/5$ there is no big difference to $\lambda = 0$. Only the $\tilde{I}$'s have a tendency to be slightly larger, which is not at all surprising. For $n = 50$ and $\lambda = 1/3$, censoring effects are more substantial. Few confidence intervals and values of $\hat{S}_n^{1\varphi}$ are somewhat large. Similarly for $\lambda = 1/2$ and $n = 50$ and $n = 100$.

The same effects will also appear in the non-modified Kaplan-Meier integral $\hat{S}_n^{\varphi}$. In view of the negative bias this seems a little unexpected. A closer look at $\hat{S}_n^{\varphi}$ (and the data) reveals the following interesting fact: $\hat{S}_n^{\varphi}$ takes on relatively large values if most of the large data are censored but the largest is uncensored.

Under heavy censoring this may happen, of course, in rare cases. It becomes a rule rather than an exception if we just replace $\hat{F}_n$ by F_n^*, i.e., if by definition we always set $\delta_{[n:n]} \equiv 1$. Our discussion points out that under heavy censoring the mean lifetime estimator is slightly non-robust. As we have seen this kind of non-robustness is not caused by outliers but by the underlying pattern of the δ's in the extreme right tails.
A closer look at the plug-in estimator $\hat{\sigma}_n^2$ shows that relatively large values are obtained if $\delta_{[n-1:n]} = 0$ and $\delta_{[n:n]} = 1$, i.e., in situations already discussed in connection with (2.4), (5.1) and(5.2). In comparison the modified Jackknife estimate of variance is much more stable. See Stute (1994d) for further details.

Figure 1: No censorship $n = 50$ $\varphi(x) = x$

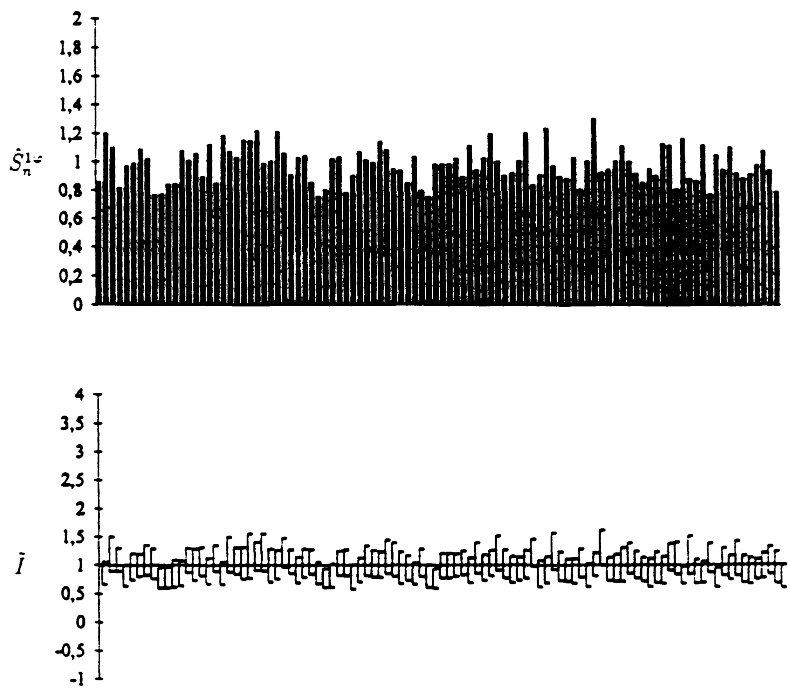


Figure 2: $\lambda = 1/5$ $n = 50$ $\varphi(x) = x$

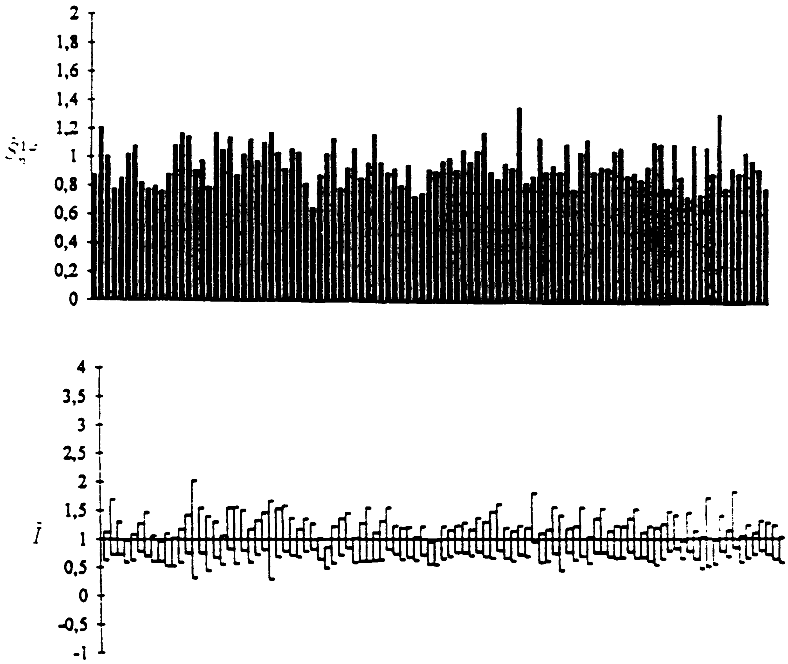


Figure 3: $\lambda = 1/3$ $n = 50$ $\varphi(x) = x$

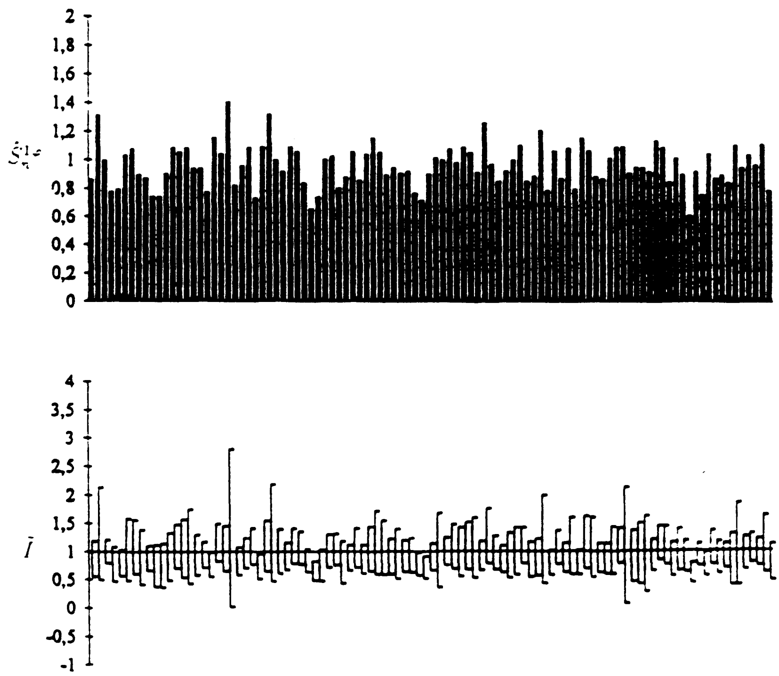


Figure 4: $\lambda = 1/2 \qquad n = 50 \qquad \varphi(x) = x$

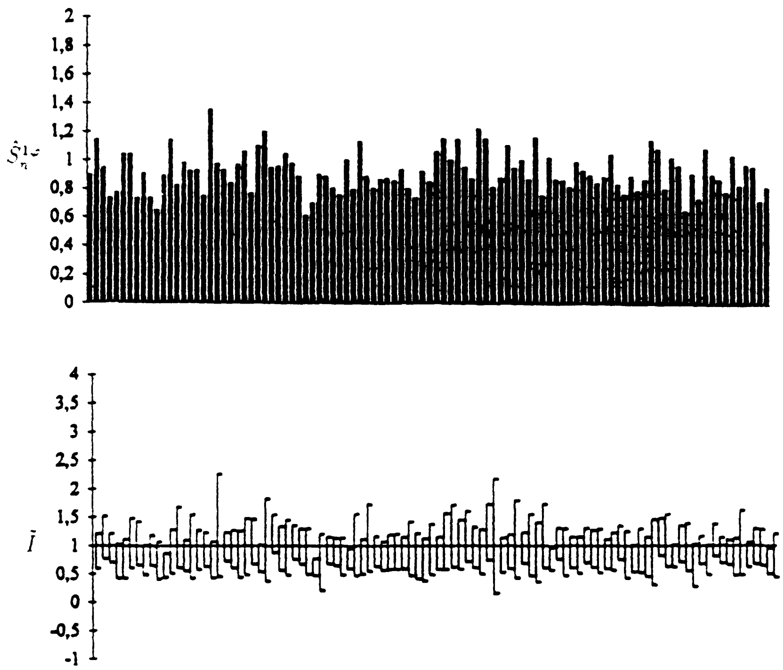
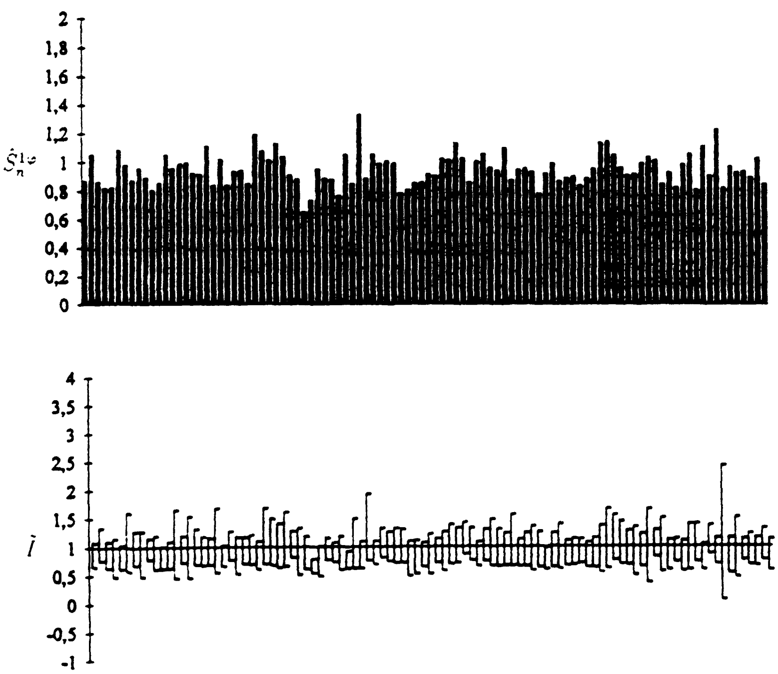


Figure 5: $\lambda = 1/2 \qquad n = 100 \qquad \varphi(x) = x$



REFERENCES

ANDERSEN, P.K., BORGAN, O., GILL, R.D. AND KEIDING, N. (1993). *Statistical models based on counting processes.* Springer, New York - Berlin.

BRESLOW, N.E. and CROWLEY, J.J. (1974). A large sample study of the life table and product-limit estimates under random censorship. *Ann. Statist.* **2** 437-453.

BURKE, M.D., CSÖRGŐ, S. AND HORVÁTH, L. (1981). Strong approximations of some biometric estimates under random censorship. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **56** 87-112.

BURKE, M.D., CSÖRGŐ, S. AND HORVÁTH, L. (1988). A correction to and improvement of 'Strong approximations of some biometric estimates under random censorship'. *Probab. Theory Rel. Fields* **79** 51-57.

CHEN, Y.Y., HOLLANDER, M. AND LANGBERG, N.A. (1982). Small sample results for the Kaplan-Meier estimator. *JASA* **77** 141-144.

EFRON, B. and TIBSHIRANI, R.J. (1993). *An introduction to the bootstrap.* Chapman and Hall, New York - London.

ETEMADI, N. (1981). An elementary proof of the strong law of large numbers. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **55** 119-122.

FLEMING, T.R. AND HARRINGTON, D.P. (1991). *Counting processes and survival analysis.* Wiley, New York.

GAVER, N.P. AND MILLER, R.G. (1983). Jackknifing the Kaplan-Meier survival estimator for censored data: simulation results and asymptotic analysis. *Commun. Statist. Theory Meth.* **A 12** 1701-1718.

GILL, R.D. (1980). Censoring and Stochastic Integrals. Math. Centre Tracts **124**. Math. Centrum, Amsterdam.

GILL, R.D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49-58.

GILL, R.D. (1994). Lectures on survival analysis. In: *Lecture Notes Math.* **1581** (Bakry, ed.) 115-241, Springer, New York - Berlin.

GRAY, H.L. and SCHUCANY, W.R. (1972). *The generalized jackknife statistics.* Dekker, New York.

GREENWOOD, M. (1926). The natural duration of cancer. *Reports on Public Health and Medical Subjects No.* **33**. His Majesty's Stationary Office.

IRWIN, J.O. (1949). The standard error of an estimate of expectational life. *J. of Hygiene* **47** 188-189.

KAPLAN, E.L. AND MEIER, P. (1958). Nonparametric estimation from incomplete observations. *JASA* **53** 457-481.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent R.V.'s and the sample DF. I. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32** 111-131.

LO, S.H. AND SINGH, K. (1986). The product-limit estimator and the bootstrap: Some asymptotic representations. *Probab. Theory Rel. Fields* **71** 455-465.

MAJOR, P. and REJTŐ, L. (1988). Strong embedding of the estimator of the distribution function under random censorship. *Ann. Statist.* **16** 1113-1132.

MAURO, D. (1985). A combinatoric approach to the Kaplan-Meier estimator. *Ann. Statist.* **13** 142-149.

NEVEU, J. (1975). *Discrete-Parameter martingales.* North-Holland, Amsterdam.

QUENOUILLE, M.H. (1956). Notes on bias in estimation. *Biometrika* **43** 353-360.

SANDER, J.M. (1975). Asymptotic normality of linear combinations of functions of order statistics with censored data. *Tech. Report No.* **8** Div. Biostatistics, Stanford University.

SCHICK, A., SUSARLA, V. and KOUL, H. (1988). Efficient estimation of functionals with censored data. *Statist. and Decisions* **6**, 349-360.

SHORACK, G.R. and WELLNER, J.A. (1986). *Empirical processes with applications to statistics.* Wiley, New York.

STUTE, W. (1994a). Strong and weak representations of cumulative hazard function and Kaplan-Meier estimators on increasing sets. *J. Statist. Planning and Inf.* **42** 315-329.

STUTE, W. (1994b). Improved estimation under random censorship. *Commun. Statist. Theory Meth.* **23** 2671-2682.

STUTE, W. (1994c). The bias of Kaplan-Meier integrals. *Scand. J. Statist.* **21** 475-484.

STUTE, W. (1994d). *The jackknife estimate of variance of a Kaplan-Meier integral.* Preprint, Univ. of Giessen.

STUTE, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23** 422-439.

STUTE, W. and WANG, J.L. (1993). The strong law under random censorship. *Ann. Statist.* **21** 1591-1607.

STUTE, W. and WANG, J.L. (1994). The jackknife estimate of a Kaplan-Meier integral. *Biometrika* **81** 602-606.

SUSARLA, V. and VAN RYZIN, J. (1980). Large sample theory for an estimator of the mean survival time from censored samples. *Ann. Statist.* **8** 1002-1016.

TUKEY, J.W. (1958). Bias and confidence in not quite large samples (Abstract). *Ann. Math. Statist.* **29** 614.

WELLNER, J.A. (1985). A heavy censoring limit theorem for the product limit estimator. *Ann. Statist.* **13** 150-162.

YANG, S. (1994). A central limit theorem for functionals of the Kaplan-Meier estimator. *Statist. Probab. Letters* **21** 337-345.

YING, Z. (1989). A note on the asymptotic properties of the product-limit estimator on the whole line. *Statist. Probab. Letters* **7** 311-314.

ZHOU, M. (1988). Two-sided bias bound of the Kaplan-Meier estimator. *Probab. Theory Rel. Fields* **79** 165-173.