NONPARAMETRIC ESTIMATION OF THE SURVIVAL DISTRIBUTION IN CENSORED DATA

by

Thomas R. Fleming and David P. Harrington

Technical Report Series, No. 8
August 1979

ABSTRACT

A nonparametric estimator of the survival distribution is presented. The estimator, applicable when dealing with censored data and related to the Kaplan-Meier estimator, is shown to be asymptotically unbiased, to be uniformly strongly consistent, and when properly normalized to converge weakly to a specified Gaussian process. Further, in small or moderate samples with untied data, the estimator is found to nearly always have smaller mean squared error than the Kaplan-Meier estimator whenever the true survival probability is at least 0.20.

KEY WORDS: Kaplan-Meier estimator; Cumulative intensity function; Mean squared error; Nonparametric estimation; Survival distribution; Censored data.

Thomas R. Fleming is an Assistant Professor in the Department of Medical Research Statistics and Epidemiology, Mayo Clinic, Rochester, MN 55901. David P. Harrington is an Assistant Professor in the Department of Applied Mathematics and Computer Science, University of Virginia, Charlottesville, VA 22901. The authors would like to thank Terry M. Therneau for his extensive assistance in generating the computer simulations. The research of the first author was supported in part by NIH grant CA 24089-01. The second author was partially supported by NSF grant ENG 78-06733.

1. INTRODUCTION AND SUMMARY

For several hundred years statisticians have attempted to formulate methods which could be used in the estimation of the survival function, $S(\cdot)$, when dealing with randomly censored data; that is, data which is incompletely observed due to causes such as early termination of a study, loss-to-follow-up, or competing risks.

Let X be a random variable denoting the time to a well-defined event, hereafter referred to as "death." The survival and intensity (or hazard rate) functions are then defined to be $S(t) = P(X > t)$ and $\nu(t) = - d/dt \ \ell n \ S(t)$ respectively, for $t > 0$, where $\nu(t)$ is assumed to be continuous.

Several authors have investigated parametric estimators for $S(t)$. Unfortunately unless the parametric assumptions can be justified from prior information, the resulting class of distribution functions satisfying these assumptions is usually not sufficiently broad. Others, such as Chiang (1968), have investigated estimators which arise when parametric assumptions are made only over subintervals. In general, estimation of conditional probabilities of survival over pre-specified sub-intervals leads to a class of estimators called "actuarial". In addition to Chiang, Berkson and Gage (1950), Elveback (1958), Gehan (1968), Cutler and Ederer (1958), and Littell(1952) have proposed well-known actuarial estimators.

More recently, sophisticated covariate analysis techniques have been developed enabling one to make use of concomitant information. Commonly the effect upon survival of such information has been modeled by means of regression on the intensity function (Cox 1972).

However, despite the useful covariate techniques, the standard nonparametric survival curve estimator remains, in practice, a most valuable, reliable, and frequently used descriptive tool. Undoubtedly, the product-limit estimator proposed by Kaplan and Meier is the most commonly employed. Assume one observes N individuals and let $\{T_j \mid 1 \leq j \leq d\}$ represent the ordered d distinct observed death times. (Here, as well as in Sections 2 through 4, we assume no ties exist among observed death times.) If the left continuous process $N(t)$ represents the number of individuals still alive and under observation just prior to t, then the Kaplan-Meier estimator is defined by

$$S_1(t) = \prod_{\{j:T_j \leq t\}} \left(1 - [N(T_j)]^{-1}\right).$$

This estimator has been shown to maximize the likelihood of the observations in the class of all possible distributions (Kaplan and Meier 1958). As $N \to \infty$, its bias converges to zero exponentially and unlike the actuarial estimators, it is strongly consistent (Aalen 1978). Further it has been shown $\sqrt{N}\,(S_1(\cdot) - S(\cdot))$ converges weakly to a mean zero Gaussian process with covariance kernel

$$S(s)\, S(t) \int_0^s [C(u)]^{-1}\, d[S(u)]^{-1}$$

for $s \leq t$, where $C(u)$ is the probability an individual is uncensored at time u (Breslow and Crowley 1974).

One somewhat undesirable feature of $S_1(t)$, however, is the fact that, especially when applied to censored data, it tends to be more variable than other estimators over the region where $N(t)$ is small. In fact, if the last observed individual dies at time γ, $S_1(t) = 0$ for all $t \geq \gamma$ independent of the value of S_1 just prior to γ. An estimator closely related to the Kaplan-Meier, but not as volatile at small $N(t)$, will be formulated below.

Let

$$\beta(t) = \int_0^t \nu(s)ds$$

be the cumulative intensity function and define its estimator by

$$\hat{\beta}(t) = \sum_{\{j:T_j \leq t\}} [N(T_j)]^{-1}$$

(see Nelson 1969). Since $\nu(t) = - d/dt\, \ell n\, S(t)$, it follows that

$$S(t) = 1 - \int_0^t S(s)d\beta(s), \tag{1.1}$$

which has solution

$$S(t) = e^{-\beta(t)}. \tag{1.2}$$

Inserting $\hat{\beta}(s)$ into (1.1) one can recursively define the estimator

$$\hat{S}(t) = 1 - \int_0^t \hat{S}(s-0)d\hat{\beta}(s), \tag{1.3}$$

where the integral is Lebesque-Stieltjes. The resulting survival estimator in (1.3) is the Kaplan-Meier. Thus there is reason to believe one may obtain an improved, at least less variable, estimator than the Kaplan-Meier by solving (1.1) first, yielding (1.2), into which one inserts the cumulative intensity estimator. The resulting estimator, $S_2(t) = e^{-\hat{\beta}(t)}$, is the one which we propose to study. Since

$$S_2(t) = \prod_{\{j:T_j \leq t\}} \exp\{-[N(T_j)]^{-1}\}, \tag{1.4}$$

$$\frac{S_1(t)}{S_2(t)} = \prod_{\{j:T_j \leq t\}} \left(\frac{1-x_j}{e^{-x_j}}\right) \tag{1.5}$$

where $x_j = [N(T_j)]^{-1}$, implying $S_1(t) \leq S_2(t)$.

Inspecting (1.4) and (1.5), $S_2(t)$ is less variable for small $N(t)$, as expected. However, since $e^{-x} \approx 1-x$ for small x, the two estimators differ little for large $N(t)$ leading one to suspect that asymptotic properties of S_1 summarized earlier can be similarly verified for S_2. This verification will be presented in Section 2 of this paper.

A small sample comparison of the mean squared error (MSE) of S_1 and S_2, of particular interest to the applied statistician, is presented in Section 3 for the situation in which the data are uncensored, and in Section 4 for censored data using Monte Carlo simulations.

It is shown in these two sections that $S_2(t) = e^{-\hat{\beta}(t)}$ has smaller MSE than the Kaplan-Meier estimator generally whenever the true survival probability $S(t)$ is greater than .20. Finally in Section 5, an application of the estimator $S_2(t)$ to data containing ties in observed death times is discussed and an example is presented.

2. THE STATISTICAL MODEL AND ASYMPTOTIC RESULTS

Assume N independent individuals are observed, each of whom is subject to arbitrary right censorship. Specifically, assume the $i^{\underline{th}}$ individual has true survival time X_i and censoring time Y_i so $S(t) = P(X_i > t)$ and $C(t) = P(Y_i > t)$. Then one is able to observe only $Z_i = \min(X_i, Y_i)$ and $\delta_i = I_{[X_i \leq Y_i]}$

where $I_{[A]}$ is the indicator random variable for the event A. Independence is assumed between the censoring and survival distributions, implying $\pi(t) \equiv P(Z_i > t) = S(t)C(t)$.

Let τ be any positive number such that $\pi(\tau) > 0$. All asymptotic results obtained in this section will be with respect to the interval $T = [0, \tau]$.

Observe that the definition of $S_2(t)$ is somewhat arbitrary for t such that $N(t) = 0$. Asymptotic results in this section are obtained by assuming

$$S_2(t) = S_2(T_d) \equiv \exp\{-\sum_{j=1}^{d} [N(T_j)]^{-1}\}$$

for all $t \geq T_d$. However, each result can be immediately generalized to apply as well to the estimator $\tilde{S}_2(t)$ where $\tilde{S}_2(t) = S_2(t)$ if $N(t) > 0$ and $\tilde{S}_2(t) = h(t)$ if $N(t) = 0$, where $h(t)$ is an arbitrary random variable such that $0 \leq h(t) \leq 1$.

2.1 Uniform Strong Consistency of $S_2(t)$

In Theorem 2.1 of this section, uniform strong consistency of $S_2(t)$ will be shown to follow from the validity of that same property for $\hat{\beta}(t)$.

<u>Theorem 2.1</u>: $\sup_{0 \leq t \leq \tau} (\sqrt{N}/\text{Log } N)|e^{-\hat{\beta}(t)} - e^{-\beta(t)}| \to 0$ a.s. as $N \to \infty$

<u>Proof</u>: Since $(e^x - 1) \leq xe^x$ for any $x \varepsilon (-\infty, \infty)$, $(e^{-\hat{\beta}(t)} - e^{-\beta(t)})$ $= e^{-\beta(t)} [e^{(\beta(t)-\hat{\beta}(t))} - 1] \leq e^{-\beta(t)} (\beta(t)-\hat{\beta}(t))\, e^{(\beta(t)-\hat{\beta}(t))}$ $= (\beta(t)-\hat{\beta}(t))\, e^{-\hat{\beta}(t)}$. Similarly, $-(e^{-\hat{\beta}(t)} - e^{-\beta(t)}) \leq (\hat{\beta}(t)-\beta(t))e^{-\beta(t)}$. Thus $[\beta(t)-\hat{\beta}(t)]\, e^{-\beta(t)} \leq e^{-\hat{\beta}(t)} - e^{-\beta(t)} \leq [\beta(t)-\hat{\beta}(t)]\, e^{-\hat{\beta}(t)}$. The theorem now follows from Proposition 3(i) of Aalen (1978).

2.2 Asymptotic Unbiasedness of $S_2(t)$

Unlike the bias of the Kaplan-Meier estimator which converges to zero exponentially, the bias of $S_2(t)$ converges to zero at a linear rate as $N \to \infty$, as shown in the following theorem.

Theorem 2.2:

$$N(E\, S_2(t) - S(t)) \to 1/2\, S(t) \int_0^t [C(s)]^{-1}\, d[S(s)]^{-1}, \text{ as } N \to \infty$$

Proof: For notational simplicity denote $N(t+0)$ by $\eta(t)$. Let $R(t) = [\eta(t)]^{-1}$ if $\eta(t) > 0$ and be 0 otherwise. Then using a technique of Aalen (1978),

$$e^{-\hat{\beta}(s+h)} = e^{-\hat{\beta}(s)}[1-I(s,h)(1-e^{-R(s)})]\ (1-U(s,h))$$

where $I(s,h)$ is an indicator of the event that at least one observed death occurs during the interval $(s, s+h]$, and the random variable $U(s,h) \in [0,1]$ such that $P(U(s,h) \neq 0) = o(h)$.

Thus, if we set $f(s) = E\, e^{-\hat{\beta}(s)}$,

$$f(s+h) - f(s) = -E\,\{I(s,h)\, e^{-\hat{\beta}(s)}\ (1-e^{-R(s)})\} + o(h).$$

Since $E(I(s,h)|\eta(s)) = \eta(s)\, \nu(s)\, h + o(h)$,
$f(s+h) - f(s) = -h\, \nu(s)\, E\{\eta(s)e^{-\hat{\beta}(s)}\ (1-e^{-R(s)})\} + o(h)$, from which we can deduce

$$\begin{aligned} f'(s) &= -\nu(s)\, E\{\eta(s)e^{-\hat{\beta}(s)}\ (1-e^{-R(s)})\} \\ &= -\nu(s)\, E\{e^{-\hat{\beta}(s)}\ I_{[\eta(s)>0]}\} + \nu(s)\, E\{\eta(s)e^{-\hat{\beta}(s)}\ [R(s) - (1-e^{-R(s)})]\} \\ &= -\nu(s)\, f(s) + \nu(s)\, E\{e^{-\hat{\beta}(s)}\ I_{[\eta(s)=0]}\} + \nu(s)\, E\{\eta(s)e^{-\hat{\beta}(s)}[R(s) - (1-e^{-R(s)})]\}. \end{aligned}$$

Adding $\nu(s)\, f(s)$ to both sides, multiplying by $e^{-\int_s^t \nu(u)du}$, and integrating, we have

$$E\, e^{-\hat{\beta}(t)} - e^{-\beta(t)} = \int_0^t \frac{d}{ds}\left[e^{-\int_s^t \nu(u)du} f(s)\right]ds = K_N(t) + L_N(t)$$

where

$$K_N(t) = \int_0^t \nu(s)\, E[I_{[\eta(s)=0]}\, e^{-\hat{\beta}(s)}]\, e^{-\int_s^t \nu(u)du}\, ds$$

and

$$L_N(t) = \int_0^t \nu(s)\, E\{\eta(s)\, e^{-\hat{\beta}(s)} \sum_{n=2}^{\infty} (-1)^n [R(s)]^n/n!\}\, e^{-\int_s^t \nu(u)du}\, ds \,.$$

$K_N(t)$, proven by Aalen to be the bias of the Kaplan-Meier estimator, clearly converges to 0 at an exponential rate.

Now

$$\left|N\, L_N(t) - \frac{S(t)}{2} \int_0^t [C(s)]^{-1}\, d[S(s)]^{-1}\right|$$

$$= \left| \int_0^t E\{e^{-\hat{\beta}(s)}\, \eta(s)\, R(s)\, N \sum_{n=2}^{\infty} (-1)^n [R(s)]^{n-1}/n!\}\, \nu(s)\, e^{-\int_s^t \nu(u)du}\, ds \right.$$

$$\left. - \frac{1}{2} \int^t e^{-\beta(s)}\, [C(s)]^{-1}\, e^{-\int_s^t \nu(u)du}\, de^{\beta(s)} \right|$$

$$= \left| \int_0^t E\{e^{-\hat{\beta}(s)}\, I_{[\eta(s)>0]}\, N[\tfrac{1}{2} R(s) - \sum_{n=3}^{\infty} \frac{[-R(s)]^{n-1}}{n!}]\}\, \nu(s)\, e^{-\int_s^t \nu(u)du}\, ds \right.$$

$$\left. - \frac{1}{2} \int_0^t e^{-\beta(s)}\, [\pi(s)]^{-1}\, \nu(s)\, e^{-\int_s^t \nu(u)du}\, ds \right|$$

$$= \Big| \frac{1}{2} \int_0^t E\{e^{-\hat{\beta}(s)} I_{[\eta(s)>0]} (NR(s) - [\pi(s)]^{-1})$$

$$+ (e^{-\hat{\beta}(s)} I_{[\eta(s)>0]} - e^{-\beta(s)}) [\pi(s)]^{-1}\} \nu(s) e^{-\int_s^t \nu(u)du} ds$$

$$- \int_0^t E\{e^{-\hat{\beta}(s)} I_{[\eta(s)>0]} N \sum_{n=3}^{\infty} \frac{[-R(s)]^{n-1}}{n!}\} \nu(s) e^{-\int_s^t \nu(u)du} ds\Big|$$

$$\leq \frac{1}{2} \{\sup_{0<s<t} E|e^{-\hat{\beta}(s)} I_{[\eta(s)>0]} (NR(s) - [\pi(s)]^{-1})|$$

$$+ \sup_{0<s<t} E|e^{-\hat{\beta}(s)} I_{[\eta(s)>0]} - e^{-\beta(s)}|[\pi(t)]^{-1}\} \int_0^t \nu(s) e^{-\int_s^t \nu(u)du} ds$$

$$+ \int_0^t E\{e^{-\hat{\beta}(s)} I_{[\eta(s)>0]} N[R(s)]^2/6\} \nu(s) e^{-\int_s^t \nu(u)du} ds$$

$$\leq \frac{1}{2} \{\sup_{0<s<t} E|NR(s) - [\pi(s)]^{-1}| + E \sup_{0<s<t} |e^{-\hat{\beta}(s)} I_{[\eta(s)>0]}$$

$$- e^{-\beta(s)}|[\pi(t)]^{-1}\} (1-S(t)) + \frac{1}{6} \sup_{0<s<t} E\, N[R(s)]^2 (1-S(t))$$

The theorem now follows by Lemma 4.2 (ii) of Aalen (1976), by Theorem 2.1, and by the fact that $P(\eta(t) = 0) \to 0$.

2.3 Weak Convergence

Let $D(T)$ be the space of functions on the interval $T = [0,\tau]$ which have discontinuities of only the first kind, and let d_0 be the Skorohod metric on $D(T)$. In this section, the term "weak convergence" will be used with respect to d_0 on $D(T)$, and will be denoted by $\Rightarrow$.

Let W represent a Wiener process and define the non-negative deterministic function g on T by $g^2(s) = [\pi(s)]^{-1} \nu(s)$. The process $\psi = \{\psi(t) | t \varepsilon T\}$ defined by

$$\psi(t) = e^{-\beta(t)} \int_0^t g(s) d\, W(s)$$

is then a mean zero Gaussian process with covariance kernel

$$S(s)\ S(t) \int_0^s [C(u)]^{-1}\ d[S(u)]^{-1} \text{ for } s \leq t$$

In Theorem 2.3, the sequence of processes $\psi_N = \{\sqrt{N}\,(e^{-\hat{\beta}(t)} - e^{-\beta(t)}) | t\varepsilon T\}$ is shown to converge weakly to the same limiting process to which $\{\sqrt{N}\,(S_1(t) - S(t)) | t\varepsilon T\}$ converges.

<u>Theorem 2.3</u>: As $N \to \infty$

$$\psi_N \equiv \{\sqrt{N}\,(S_2(t) - S(t)) | t\varepsilon T\} \Rightarrow \psi \equiv \{e^{-\beta(t)} \int_0^t g(s) dW(s) | t\varepsilon T\}.$$

<u>Proof</u>: By Taylor's series expansion,

$$\sqrt{N}\,(S_2(t) - S(t)) = \sqrt{N}\, e^{-\beta(t)}\ [e^{(\beta(t)-\hat{\beta}(t))} - 1] \tag{2.1}$$

$$= \sqrt{N}\, e^{-\beta(t)}\ \{(\beta(t)-\hat{\beta}(t)) + (\beta(t)-\hat{\beta}(t))^2 \sum_{n=0}^{\infty} [\frac{(\beta(t)-\hat{\beta}(t))^n}{(n+2)!}]\}$$

Observe that

$$\sup_{0\leq t\leq \tau} |[\beta(t)-\hat{\beta}(t)] \sum_{n=0}^{\infty} \frac{[\beta(t)-\hat{\beta}(t)]^n}{(n+2)!}| \tag{2.2}$$

$$\leq \sup_{0\leq t\leq \tau} |\beta(t)-\hat{\beta}(t)| e^{|\beta(t)-\hat{\beta}(t)|} \to 0 \text{ a.s.}$$

by Proposition 3(i) of Aalen (1978).

Further, by Theorem 8.2 of Aalen (1976),

$$\{\sqrt{N}(\beta(t)-\hat{\beta}(t))\,|\,t\varepsilon T\} \Rightarrow \{\int_0^t g(s)dW(s)\,|\,t\varepsilon T\}. \tag{2.3}$$

Using Corollary 1 of Theorem 5.1 from Billingsley (1968), (2.2) and (2.3) imply

$$\{\sqrt{N}\, e^{-\beta(t)}\, (\beta(t)-\hat{\beta}(t))^2 \sum_{n=0}^{\infty} \frac{[\beta(t)-\hat{\beta}(t)]^n}{(n+2)!}\,|\,t\varepsilon T\} \Rightarrow 0. \tag{2.4}$$

Since (2.3) implies

$$\{\sqrt{N}\, e^{-\beta(t)}\, (\beta(t)-\hat{\beta}(t))\,|\,t\varepsilon T\} \Rightarrow \{e^{-\beta(t)} \int_0^t g(s)dW(s)\,|\,t\varepsilon T\} \tag{2.5}$$

one can see from (2.1) that the Theorem follows from (2.4) and (2.5) by once again applying Corollary 1 of Theorem 5.1 of Billingsley (1968).

3. MEAN SQUARED ERROR IN SMALL SAMPLES OF UNCENSORED DATA

Since mean squared error is generally accepted to be a primary measure of accuracy of an estimator we examined in small samples the behavior of the MSE of both S_1 and S_2 as a function of $S(t)$. In this section, as in Section 4 when investigating MSE in censored data, it is assumed no ties exist in the data. To evaluate MSE we initially defined

$$S_1(t) = S_1(T_d) \equiv \prod_{j=1}^{d} \{1-[N(T_j)]^{-1}\}$$

and

$$S_2(t) = S_2(T_d) \equiv \exp\{-\sum_{j=1}^{d} [N(T_j)]^{-1}\}$$

for all $t \geq T_d$. Suppose $p = S(t)$, and let $\tilde{\mu}_1(p,N) = E(S_1(t) - p)^2$ and $\tilde{\mu}_2(p,N) = E(S_2(t) - p)^2$. In uncensored data, it is well known that $\tilde{\mu}_1(p,N) = p(1-p)/N$. Further, if one takes $S_k = 1 + 1/2 + 1/3 + \ldots + 1/k$ it is not hard to show that

$$\tilde{\mu}_2(p,N) = e^{-2S_N} \left[\left(e^{S_N} - e^{S_N} p\right)^2 p^N + \left(e^{S_{N-1}} - e^{S_N} p\right)^2 \binom{N}{N-1} p^{N-1} (1-p) + \ldots + (1 - e^{S_N} p)^2 N(1-p)^N\right].$$

Values of $\tilde{\mu}_1(p,N)$, $\tilde{\mu}_2(p,N)$ and $\tilde{r}(p,N) \equiv \tilde{\mu}_2(p,N)/\tilde{\mu}_1(p,N)$ have been computed for $p = .02u$, $u = 0, 1, 2, \ldots, 50$, for each of the values $N = 2, 3, 4, 5, 6, 8, 10, 15, 20, 30, 40, 50$, and 100. Some representative results are given in Table 1.

The table shows that when $S(t) \approx 0$, the Kaplan-Meier estimator (in this case the empirical distribution function estimator) is substantially more accurate, but that when $S(t) \gtrapprox .2$, S_2 is more efficient in the MSE sense.

An alternative and probably preferable method for evaluating the relative accuracy, i.e. relative MSE, of $S_2(t)$ to $S_1(t)$ will now be described. Since

$$\hat{\beta}(t) = \sum_{\{j:T_j \leq t\}} [N(T_j)]^{-1},$$

we have that

$$\hat{\beta}(t) \leq \sum_{k=1}^{N} (N-k+1)^{-1} \equiv b,$$

and $\lim_{t \to \infty} S_2(t) = e^{-b} > 0$. In fact, when the data are censored $S_1(t)$ and $S_2(t)$ both have a nonzero limit with positive probability, with that limit dependent upon the observed censoring pattern. Let $\gamma = \inf \{t:N(t) = 0\}$. It seems that a more natural way to evaluate the accuracy of $S_2(t)$ might be to observe that if $N(t) = 0$, i.e. if $t > \gamma$, $S_2(t)$ does not provide a point estimate, but rather an interval estimate assigning uniform likelihood over $[0,e^{-b}]$. With this interpretation, the squared error of $S_2(t)$, conditional on $N(t) = 0$, would be

$$e^b \int_0^{e^{-b}} (u-S(t))^2 du = 1/3\ e^{-2b} - S(t)\ e^{-b} + [S(t)]^2.$$

When the formulas yielding the values in Table 1 are suitably modified, the mean squared errors now denoted by $\mu_1(p,N)$ and $\mu_2(p,N)$ and their ratio $r(p,N)$ are obtained. Since $\mu_1(p,N) = \tilde{\mu}_1(p,N)$ is tabulated in Table 1, only values of $r(p,N)$ appear in Table 2.

It is clear by comparing Tables 1 and 2 that the new interpretation of the information provided by $S_2(t)$ for $t > \gamma$ does not make a substantial change in the assessment of its efficiency relative to $S_1(t)$. This result could be anticipated since, except at values of t such that $S(t) \approx 0$, the probability that $N(t)$ is positive, i.e. that $t < \gamma$, is nearly 1 even for small sample sizes.

It can be inferred that, when applied to small or moderate sized samples of uncensored data, $S_2(t) = e^{-\hat{\beta}(t)}$ has smaller average squared deviation from the true value $S(t)$ than does the Kaplan-Meier or empirical distribution function when the true value $S(t) \geq 0.2$.

4. MEAN SQUARED ERROR IN SMALL SAMPLES OF UNCENSORED DATA

4.1 Method

It is possible to calculate relative efficiencies, i.e., relative MSE, of S_2 and S_1 similar to those discussed in Section 3 but more generally now for the situation in which the data are subject to various patterns of censorship. As noted earlier, these efficiencies may be calculated by assuming that S_1 and S_2 remain constant after γ, the last observation time, or by assuming both provide interval estimates for $S(t)$ after that last observation. Since one can anticipate either approach would lead to the same relative efficiency, only the latter was employed in this Section.

Assuming, without loss of generality, that the true survival distribution is U(0,1), i.e. $S(t) = 1-t$ for $0 \le t \le 1$, three different censoring distributions yielding mild, moderate or severe censorship were considered. Specifically, C(t) was assumed to be U(0,2) yielding 25% censorship, then U(0,1) yielding 50% censorship, and finally U(0,1/2) yielding 75% censorship. Samples of size N = 10, 20, 35, 50, 100 were considered.

Since the exact formulas for $\mu_1(p,N)$ and $\mu_2(p,N)$ are very difficult to obtain in censored data even for N as small as 5, Monte Carlo simulations were employed to obtain approximations.

For each N and each censoring distribution C, an independent collection of N pairs (X_i,Y_i) were randomly generated such that the independent random variables X_i and Y_i were distributed as S and C respectively. After determining (Z_i, δ_i) for $i = 1, \ldots, N$, S_1 and S_2 were calculated. The steps above were repeated n = 5000 times where, for the j^{th} repetition, the estimators are denoted S_{1j} and S_{2j}. The estimator S_{ij} and its squared error M_{ij} were recorded at $S(t) \equiv p = .02u$, $u = 0, \ldots, 50$, where as before

$$M_{ij}(t) = \begin{cases} [S_{ij}(t) - S(t)]^2 & \text{if } t \le \gamma \\ b_{ij}^{-1} \int_0^{b_{ij}} (u-S(t))^2\, du & \text{if } t > \gamma \end{cases}$$

with

$$b_{1j} \equiv S_1(T_{d_j}) = \prod_{k=1}^{d_j} \{1-[N(T_{jk})]^{-1}\}$$

$$b_{2j} \equiv S_2(T_{d_j}) = \exp\{\sum_{k=1}^{d_j} [N(T_{jk})]^{-1}\},$$

and $\{T_{jk} | 1 \le k \le d_j\}$ representing the set of d_j distinct observed death times in the j^{th} repetition.

If $C(t) = 1-t/a$ for $0 \le t \le a$, then the mean squared error of $S_{ij}(t)$ should be denoted by $EM_{ij}(t) = \mu_i(p,N,a)$ since it is now a function not only of p and N, but of the censoring pattern relative to the survival distribution as well. However, in the following notation p, N and a will be deleted.

If we define $\sigma_i^2 = E(M_{ij} - \mu_i)^2$, then the coefficient of variation of M_{ij} is

$$\sigma_i/\mu_i = \sqrt{\frac{EM_{ij}^2}{(EM_{ij})^2} - 1},$$

which incidentally can be shown to converge to $\sqrt{2}$ as sample size N grows large.

Observe that $\{(M_{1j}, M_{2j}): j = 1, \ldots, n\}$ is an independent collection of bivariate random variables such that $EM_{kj} = \mu_k$ and $E[(M_{kj} - \mu_k)(M_{\ell j} - \mu_\ell)] \equiv \sigma_{k\ell}$, where we have previously denoted $\sigma_{kk} = \sigma_k^2; k, \ell \varepsilon \{1,2\}$.

Define the obvious estimator of μ_i by

$$\hat{\mu}_i = \bar{M}_i = 1/n \sum_{j=1}^{n} M_{ij}.$$

The relative MSE, $r = \mu_2/\mu_1$, will then be estimated by $\hat{r} = \hat{\mu}_2/\hat{\mu}_1 = \bar{M}_2/\bar{M}_1$. Observe

$$E\bar{M}_i = \mu_i, \text{ var } \bar{M}_i = 1/n\, \sigma_i^2 \text{ and } \text{cov}(\bar{M}_1, \bar{M}_2) = 1/n\, \sigma_{12}. \tag{4.1}$$

In order to determine var $\hat{r} \equiv \sigma_r^2$, define $f(x_1, x_2) = x_2/x_1$. Let $i, j \varepsilon \{1,2\}$. By Taylor's series expansion, if $f_i = \partial f/\partial x_i$ and $f_{ij} = \partial^2 f/\partial x_i \partial x_j$, then

$$f(\bar{M}_1, \bar{M}_2) = f(\mu_1, \mu_2) + \sum_{i=1}^{2} f_i(\mu_1, \mu_2)(\bar{M}_i - \mu_i)$$

$$+ 1/2 \sum_{i=1}^{2} \sum_{j=1}^{2} f_{ij}(\alpha_1, \alpha_2)(\bar{M}_i - \mu_i)(\bar{M}_j - \mu_j)$$

for some α_i between μ_i and $\bar{M}_i$.

Thus

$$\begin{aligned} \bar{M}_2/\bar{M}_1 - \mu_2/\mu_1 &= \frac{\bar{M}_2 - \mu_2}{\mu_1} - \frac{\mu_2}{\mu_1^2}(\bar{M}_1 - \mu_1) \\ &- \frac{(\bar{M}_2 - \mu_2)(\bar{M}_1 - \mu_1)}{\alpha_1^2} + \frac{\alpha_2 (\bar{M}_1 - \mu_1)^2}{\alpha_1^3} . \end{aligned} \tag{4.2}$$

By (4.2), one can show

$$E\, \bar{M}_2/\bar{M}_1 - \mu_2/\mu_1 = 0(1/n) \tag{4.3}$$

where we denote $g_n = 0(n^{-\alpha})$ if $\lim_{n \to \infty} n^{\alpha} g_n = k$ for some constant k.

Further, by (4.2), it can be shown that

$$\begin{aligned} E(\bar{M}_2/\bar{M}_1 - \mu_2/\mu_1)^2 &= \frac{E(\bar{M}_2 - \mu_2)^2}{\mu_1^2} + \frac{\mu_2^2}{\mu_1^4} E(\bar{M}_1 - \mu_1)^2 \\ &- \frac{2\mu_2}{\mu_1^3} E\left[(\bar{M}_1 - \mu_1)(\bar{M}_2 - \mu_2)\right] + 0(1/n^2). \end{aligned} \tag{4.4}$$

Using (4.1), (4.3) and (4.4)

$$\sigma_r^{\,2} = E\left(\frac{\overline{M}_2}{\overline{M}_1} - E\,\frac{\overline{M}_2}{\overline{M}_1}\right)^2 = \frac{1}{n}\left(\frac{\mu_2}{\mu_1}\right)^2\left(\frac{\sigma_2^{\,2}}{\mu_2^{\,2}} - 2\,\frac{\sigma_{12}}{\mu_1\mu_2} + \frac{\sigma^2}{\mu_1^{\,2}}\right) + 0\left(\frac{1}{n^2}\right)$$

Thus for large n, the coefficient of variation of $\hat{r} = \overline{M}_2/\overline{M}_1$ is approximately

$$\frac{1}{\sqrt{n}}\sqrt{\frac{\sigma_2^{\,2}}{\mu_2^{\,2}} - 2\,\frac{\sigma_{12}}{\mu_1\mu_2} + \frac{\sigma_1^{\,2}}{\mu_1^{\,2}}}$$

Its obvious estimator is given by

$$\hat{\sigma}_r/\hat{r} = \frac{1}{\sqrt{n}}\sqrt{\frac{\hat{\sigma}_2^{\,2}}{\hat{\mu}_2^{\,2}} - 2\,\frac{\hat{\sigma}_{12}}{\hat{\mu}_1\hat{\mu}_2} + \frac{\hat{\sigma}_1^{\,2}}{\hat{\mu}_1^{\,2}}}$$

where

$$\hat{\sigma}_{k\ell} = \frac{1}{n}\sum_{j=1}^{n}(M_{kj}M_{\ell j}) - \overline{M}_k\overline{M}_\ell \quad \text{for } k,\ell \in \{1,2\}.$$

4.2 Results

Table 3 contains a representative selection of the results obtained from the Monte Carlo simulations of data 75%, 50% or 25% censored. For each censoring distribution the table presents $\hat{\mu}_i$, the estimated MSE of S_i, multiplied by 1000 at times at which the true survival probability $S(t) = p$ is .9, .7, .5, .3 or .1. Sample sizes N equal to 10, 20, 35, 50 or 100 are considered.

The coefficient of variation of $\hat{\mu}_i$, not presented in Table 3, was found to be approximately equal to $\sqrt{2}/\sqrt{5000} = .02$ in most cases, as predicted earlier for large values of N. In addition to values of $\hat{\mu}_i$ Table 3 also includes $\hat{r} = \hat{\mu}_2/\hat{\mu}_1$, the estimated relative MSE, as well as $\hat{\sigma}_r/\hat{r}$, the estimated coefficient of variation of $\hat{r}$, which is multiplied by 1000.

It appears from the table that for all censoring distributions inspected and sample sizes 10 to 100, the estimator $S_2(t) = e^{-\hat{\beta}(t)}$ has smaller MSE than the Kaplan-Meier estimator $S_1(t)$ when the true survival probability S(t) is at least 0.2, a result consistent with that obtained for uncensored data in Section 3. Table 4 makes this fact more obvious. Define $A = \{(.02)u : u = 0, 1, \ldots, 50\}$ and let p_ℓ represent the largest element in A such that $\hat{r} - 2\hat{\sigma}_r > 1$ at all values p in A smaller than or equal to p_ℓ. Similarly let p_u represent the smallest element of A such that $\hat{r} + 2\hat{\sigma}_r < 1$ at all values p in A greater than or equal to p_u. Then (p_ℓ, p_u) forms a type of "confidence interval" for the region or point at which $r = 1$; that is, with reasonable certitude we can state that $r > 1$ for all $p \leq p_\ell$ and $r < 1$ for all $p \geq p_u$. Table 4 contains (p_ℓ, p_u) for the various censoring distributions and sample sizes discussed earlier. It also contains for uncensored data (p_ℓ^*, p_u^*), the true region containing the point at which $r = 1$, where $p_\ell^*, p_u^* \in A$.

One result from the Monte Carlo simulations of censored data which is not shown in Tables 3 or 4 and which was unlike results obtained for uncensored data in Section 3 relates to the behavior of r when $p = S(t) \approx 0$. In Tables 1 and 2 it was observed that $r >> 1$ when $p \approx 0$ since $\mu_1 \to 0$ as $p \to 0$. However since μ_1 fails to converge to 0 as $p \to 0$ in censored data, it was found that r was not substantially greater than 1 for $p \approx 0$ when the data was moderately censored.

It is also of interest to comment briefly on the situation in which $S(t) = 1-t$, $0 \leq t \leq 1$; and $C(t) = 1-2t$, $0 \leq t \leq 1/2$. In this case one would normally be interested in the estimator of S(t) only over the region for which we have data; i.e. $0 \leq t \leq 1/2$. One can observe from Table 4 that $r = \mu_2/\mu_1$ is less than 1 throughout that region.

4.3 Conclusion

It can be concluded quite generally that, when dealing with censored or uncensored data having relatively infrequent or no ties in observed failure times, the estimator $S_2(t) = e^{-\hat{\beta}(t)}$ is a more efficient estimator of the true survival probability S(t) than the Kaplan-Meier estimator $S_1(t)$, whenever $S(t) \geq .2$. This result is particularly enticing in many survival studies, such as in biomedical research in which one or more groups of patients are often followed for a limited period of time after onset of disease or initiation of treatment, since one is quite certain that the true survival function is much larger than .2 over the entire interval on which it is being estimated.

Examples include epidemiological record studies estimating survival of men with carcinoma of the prostate following radiation treatment or radical prostatectomy, or prospectively randomized clinical trials comparing various chemotherapeutic agents in women with breast cancer.

5. TIED DATA AND AN EXAMPLE

In this section we will discuss the estimators $S_i(t)$; $i = 1, 2$; when applied to data in which ties exist between observed death times. Recall that for the N individuals under observation, we have defined $\{T_j : 1 \leq j \leq d\}$ to represent their d distinct ordered observed times of death. If we let $D(t)$ represent the number of deaths observed at t, then we are now eliminating our previous assumption that $D(T_j) = 1$ for $j = 1, \ldots, d$.

The Kaplan-Meier estimator in this situation is defined by

$$S_1(t) = \prod_{\{j : T_j \leq t\}} \left(1 - \frac{D(T_j)}{N(T_j)}\right)$$

Momentarily formulate a new data set having no ties among observed death times by assigning, to the $D(T_j)$ individuals observed to die at T_j, distinct new times of death each infinitesimally to the left of T_j. Observe that Kaplan-Meier estimator has the interesting property that, when applied to this new data set, it yields the same estimate of S that was obtained when it was applied to the original set containing tied data. It would be desirable to have S_2 share this same interesting property so that the comparative behavior of S_1 and S_2 established in previous sections would carry over to the tied data situation.

Clearly this will be the case if we define S_2 by $S_2(t) = \exp(-\hat{\beta}(t))$ where

$$\hat{\beta}(t) = \sum_{\{j:T_j \le t\}} \sum_{k=0}^{D(T_j)-1} (N(T_j) - k)^{-1}$$

The next example shows the straightforward calculation of S_1 and S_2 in a set of data containing ties.

<u>Example 1</u>. Recently a Phase II clinical trial was conducted at the Mayo Clinic to determine the efficacy of an anti-tumor chemotherapeutic agent, Maytanzine, in patients with advanced head and neck carcinoma. Survival times of the 22 patients from first day of treatment were 18, 19, 23, 23, 23*, 44, 54*, 74, 74, 96, 109*, 114*, 119*, 125*, 133, 135*, 141*, 156*, 167*, 238, 253* and 283 days, where * denotes a censored observation. Table 5 contains the calculation of S_1 and S_2 for this data. Observe that the Kaplan-Meier estimator S_1 and $S_2 = \exp(-\hat{\beta})$ differ little over the region in which $N(t)$ is more than ten. However, as expected from earlier discussion, S_1 is much more volatile when $N(t)$ is less than 5, with S_1 dropping from 0.36 to 0 at 283 days due to the one death observed at that time.

REFERENCES

Aalen, Odd (1976), "Nonparametric Inference in Connection with Multiple Decrement Models," Scandinavian Journal of Statistics, 3, 15-27.

Aalen, Odd (1978), "Non-parametric Estimation of Partial Transition Probabilities in Multiple Decrement Models," Annals of Statistics, 6, 534-545.

Berkson, Joseph and Gage, Robert (1950), "Calculation of Survival Rates for Cancer," Proceedings of the Staff Meetings of the Mayo Clinic, 25, 270-286.

Billingsley, Patrick (1968), Convergence of Probability Measures, New York: John Wiley and Sons.

Breslow, Norman and Crowley, John (1974), "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship," Annals of Statistics, 2, 437-454.

Chiang, Chin Long (1968), Introduction to Stochastic Processes in Biostatistics, New York: John Wiley and Sons.

Cox, David R. (1972), "Regression Models and Life Tables," Journal of the Royal Statistical Society, Ser. B, 34, 187-220.

Cutler, Sidney and Ederer, Fred (1958), "Maximum Utilization of the Life Table Method in Analyzing Survival," Journal of Chronic Diseases, 8, 699-712.

Elveback, Lila (1958), "Actuarial Estimation of Survivorship in Chronic Disease," Journal of the American Statistical Association, 53, 420-440.

Gehan, Edmund (1968), "Estimating Survival Functions from the Life Table," Journal of Chronic Diseases, 21, 629-644.

Kaplan, E. L. and Meier, Paul (1958), "Non-parametric Estimation from Incomplete Observations," Journal of the American Statistical Association, 53, 457-481.

Littell, Arthur S. (1952), "Estimation of the T-year Survival Rate from Follow-up Studies Over a Limited Period of Time," Human Biology, 24, 87-116.

Nelson, Wayne B. (1969), "Hazard Plotting for Incomplete Failure Data," Journal of Quality Technology, 1, 27-52.

TABLE I

$$\frac{\tilde{\mu}_2(p,N) \times 10^3 \;/\; \tilde{\mu}_1(p,N) \times 10^3}{\tilde{r}(p,N)}$$

p = S(t)	N = Sample Size: 10	20	50	100
.98	1.776/ 1.960 .906	0.933/ 0.980 .952	0.384/0.392 .980	0.194/0.196 .990
.94	5.115/ 5.640 .907	2.685/ 2.820 .952	1.106/1.128 .981	0.558/0.564 .990
.9	8.172/ 9.000 .908	4.287/ 4.500 .953	1.765/1.800 .981	0.891/0.900 .990
.8	14.576/16.000 .911	7.635/ 8.000 .954	3.141/3.200 .981	1.585/1.600 .990
.7	19.212/21.000 .915	10.042/10.500 .956	4.126/4.200 .982	2.081/2.100 .991
.6	22.077/24.000 .920	11.510/12.000 .959	4.721/4.800 .983	2.380/2.400 .991
.5	23.170/25.000 .927	12.037/12.500 .963	4.925/5.000 .985	2.481/2.500 .993
.4	22.488/24.000 .937	11.624/12.000 .969	4.740/4.800 .987	2.385/2.400 .993
.3	20.032/21.000 .954	10.286/10.500 .978	4.164/4.200 .991	2.091/2.100 .996
.2	15.831/16.000 .989	7.967/ 8.000 .996	3.198/3.200 .999	1.599/1.600 .999
.1	9.993/9.000 1.110	4.732/ 4.500 1.051	1.840/1.800 1.022	0.911/0.900 1.012
.06	7.263/ 5.640 1.288	3.200/ 2.820 1.135	1.187/1.128 1.051	0.579/0.564 1.027
.02	4.358/ 1.960 2.223	1.575/ 0.980 1.607	0.479/0.392 1.223	0.216/0.196 1.105

TABLE II

r(p,N)

p = S(t)	N = Sample size 10	20	50	100
.98	.906	.952	.980	.990
.94	.907	.952	.981	.990
.90	.908	.953	.981	.990
.80	.911	.954	.981	.991
.70	.915	.956	.982	.991
.60	.920	.959	.983	.992
.50	.928	.963	.985	.993
.40	.942	.969	.987	.994
.30	.973	.979	.991	.996
.20	1.048	1.003	.999	.999
.10	1.244	1.112	1.025	1.012
.06	1.412	1.252	1.075	1.028
.02	1.875	1.639	1.353	1.166

TABLE III

$\hat{r} = \hat{\mu}_2/\hat{\mu}_1$; i.e., MSE of $S_2(t) = e^{-\hat{\beta}(t)}$ relative to MSE of $S_1(t)$, the Kaplan-Meier estimator

$\hat{\mu}_2 \times 10^3 / \hat{\mu}_1 \times 10^3$

$\hat{r}$; $(\hat{\sigma}_r / \hat{r}) \times 10^3$

p = S(t)	N = 10	N = 20	N = 35	N = 50	N = 100
			C(t) ~ U(0, 1/2)		
.9	9.40/10.53 .8929; 2.13	4.76/ 5.04 .9456; 1.17	2.84/ 2.94 .9671; 0.87	2.00/ 2.05 .9766; 0.72	0.98/ 0.99 .9884; 0.55
.7	36.39/42.61 .8541; 4.91	16.21/17.85 .9082; 3.51	9.38/ 9.91 .9464; 2.29	6.68/ 6.92 .9651; 1.82	3.31/ 3.37 .9823; 1.27
.5	84.02/98.74 .8509; 3.66	83.50/93.76 .8905; 2.99	83.88/91.69 .9148; 2.61	83.43/89.36 .9337; 2.24	83.58/87.41 .9563; 1.81
.3	47.64/51.33 .9282; 4.60	38.56/41.04 .9396; 4.27	33.59/35.55 .9450; 4.05	31.55/32.92 .9586; 3.55	28.17/29.02 .9707; 3.06
.1	91.27/83.92 1.0876; 1.51	73.63/68.32 1.0776; 1.19	63.30/59.40 1.0656; 0.98	59.68/56.48 1.0566; 0.84	52.75/50.63 1.0418; 0.62
			C(t) ~ U(0, 1)		
.9	8.29/ 9.17 .9034; 1.72	4.43/ 4.66 .9503; 1.15	2.72/ 2.80 .9719; 0.84	1.82/ 1.86 .9802; 0.72	0.92/ 0.93 .9892; 0.54
.7	23.17/26.06 .8889; 3.85	12.13/12.89 .9413; 2.60	6.84/ 7.06 .9686; 1.83	5.10/ 5.22 .9780; 1.49	2.61/ 2.64 .9901; 1.08
.5	36.66/41.11 .8917; 5.31	18.26/19.79 .9225; 4.51	10.25/10.70 .9576; 3.23	7.22/ 7.43 .9715; 2.67	3.79/ 3.83 .9896; 1.81
.3	46.42/49.81 .9320; 5.95	25.55/27.41 .9320; 5.76	14.09/15.04 .9367; 5.58	9.09/ 9.63 .9443; 5.25	4.58/ 4.67 .9798; 3.47
.1	28.01/24.96 1.1221; 5.07	18.04/16.93 1.0653; 5.71	11.44/11.29 1.0137; 6.99	9.26/ 9.10 1.0185; 6.70	5.53/ 5.59 .9904; 6.45
			C(t) ~ U(0, 2)		
.9	8.45/ 9.32 .9065; 1.63	4.41/ 4.64 .9505; 1.09	2.56/ 2.63 .9713; 0.83	1.79/ 1.83 .9789; 0.70	0.94/ 0.96 .9892; 0.53
.7	21.49/23.80 .9028; 3.30	11.15/11.76 .9481; 2.21	6.33/ 6.50 .9738; 1.69	4.41/ 4.50 .9807; 1.43	2.31/ 2.33 .9901; 1.04
.5	28.23/30.95 .9120; 4.98	14.78/15.61 .9467; 3.59	8.37/ 8.61 .9718; 2.69	5.71/ 5.82 .9794; 2.28	3.02/ 3.05 .9913; 1.61
.3	29.62/30.19 .9812; 5.98	14.16/14.69 .9643; 5.50	8.15/ 8.38 .9721; 4.39	5.50/ 5.59 .9841; 3.67	2.72/ 2.75 .9909; 2.62
.1	20.01/16.02 1.2487; 7.04	9.21/ 7.87 1.1698; 6.80	4.65/ 4.33 1.0736; 6.79	3.22/ 3.05 1.0564; 6.28	1.46/ 1.45 1.0060; 5.45

TABLE IV

"Confidence Intervals" for Points $S(t) = p$ at which $r = 1$

N	$C(t) \sim U(0, 1/2)$ (p_ℓ, p_u)	$C(t) \sim U(0,1)$ (p_ℓ, p_u)	$C(t) \sim U(0,2)$ (p_ℓ, p_u)	Uncensored data (p_ℓ^*, p_u^*)
10	.22 - .24	.14 - .18	.24 - .30	.24 - .26
20	.22 - .26	.12 - .18	.20 - .24	.20 - .22
35	.22 - .26	.08 - .14	.14 - .20	.18 - .20
50	.22 - .26	.10 - .14	.12 - .20	.18 - .20
100	.22 - .26	.06 - .12	.08 - .24	.18 - .20

TABLE V

Calculation of S_1 and S_2 for Head and Neck Data

k	T_k	$N(T_k)$	$D(T_k)$	$\hat{B}(T_k)-\hat{B}(T_{k-1})$	$\hat{B}(T_k)$	$S_1(T_k)$	$S_2(T_k)$
0	0	22	0		0	1	1
1	18	22	1	1/22	.0455	.9545	.9556
2	19	21	1	1/21	.0931	.9091	.9111
3	23	20	2	1/20 + 1/19	.1957	.8182	.8226
4	44	17	1	1/17	.2545	.7701	.7753
5	74	15	2	1/15 + 1/14	.3926	.6674	.6753
6	96	13	1	1/13	.4695	.6160	.6253
7	133	8	1	1/8	.5945	.5390	.5518
8	238	3	1	1/3	.9279	.3594	.3954
9	283	1	1	1	1.9279	.0000	.1455