SURVIVAL ESTIMATION WITH CENSORED DATA

ALICE S. WHITTEMORE
JOSEPH B. KELLER

TECHNICAL REPORT NO. 69
APRIL 1983

PREPARED UNDER THE AUSPICES OF
SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SURVIVAL ESTIMATION WITH CENSORED DATA

BY

ALICE S. WHITTEMORE
AND
JOSEPH B. KELLER

STANFORD UNIVERSITY

TECHNICAL REPORT NO. 69

APRIL 1983

STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH (SIMS)

PREPARED UNDER SUPPORT TO SIMS FROM
ENVIRONMENTAL PROTECTION AGENCY (EPA)
SLOAN FOUNDATION
NATIONAL SCIENCE FOUNDATION (NSF)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
SURVIVAL ESTIMATION WITH CENSORED DATA

Alice S. Whittemore*
Department of Family, Community, Preventive Medicine
Stanford University Medical School
Stanford, CA 94305

and

Joseph B. Keller**
Departments of Mathematics & Mechanical Engineering
Stanford University
Stanford, CA 94305

Abstract

New nonparametric methods are given for estimating survival probability using randomly right-censored data. A class of estimators is obtained by the maximum likelihood method, in which the hazard rate is approximated by a suitably chosen spline function. The class includes the estimator proposed by Nelson (1969) and Altshuler (1970). The estimators are shown to be uniformly consistent and to have the same asymptotic weak convergence properties as the Kaplan-Meier estimator. In small and in heavily censored samples, new estimators in the class have uniformly smaller mean squared error than do the Kaplan-Meier (1958) and Nelson-Altshuler estimators, according to computer simulations. The methods are extended to provide new estimates for both the baseline hazard rate and the regression coefficients in the proportional hazards model proposed by Cox (1972). Several existing estimates for these quantities, including that obtained using Cox' partial likelihood, occur as special cases of the general procedure. The methods are illustrated using experimental carcinogenesis data.
Introduction

The estimation of survival probabilities using incomplete data has intrigued investigators for centuries. Typically one wishes to estimate the probability $F(x)$ of surviving beyond time $x$ when the (random) death times $X$ have been censored on the right. Many parametric procedures for estimating $F$ have been suggested, in which parameters in an assumed functional form for $F$ are estimated. (See Kalbfleisch and Prentice (1980) for a review.)

Kaplan and Meier (1958) proposed a simple nonparametric procedure for estimating $F$ that is widely used in medicine, life insurance and reliability theory. Nelson (1969) and Altshuler (1970) independently developed a similar procedure. Unlike the parametric estimators, these two procedures make no assumptions on $F$ other than the basic one of independence between failure and censoring variables. However they share three disadvantages. First, they do not provide a useful estimate of the hazard rate $\lambda(x) = -\frac{d}{dx} \ln F(x)$, which is often of central interest. Second, they do not use the precise censoring times. Third, their efficiency in large samples is poor. Miller (1981a) found asymptotic efficiency as low as 50% for the Kaplan-Meier estimator relative to certain parametric estimators.

We shall present a class of nonparametric estimators for $F$ that i) provide estimates for $\lambda$, ii) use the precise censoring times, and iii) have uniformly better small sample properties than do the Kaplan-Meier and Nelson-Altshuler estimators. The new estimators are obtained by the maximum likelihood method, in which $\lambda$ is replaced by a suitably chosen spline function. They are shown to have the same asymptotic weak convergence properties as do the Kaplan-Meier and Nelson-Altshuler estimators, so they provide no gain in asymptotic efficiency. However they are considerably more efficient in small and moderate samples, and their efficiency (relative to the Kaplan-Meier and Nelson-Altshuler estimators)
increases as the censoring becomes heavier.

We describe the estimators and their asymptotic properties in sections 3 and 4. Section 5 contains a comparison of the small sample properties of two of the estimators with those of the Kaplan-Meier estimator, using "data" generated by computer from various survivor functions F. In section 6 we extend the method to estimate the baseline hazard rate and regression coefficients in the proportional hazard model proposed by Cox (1972). Examples are presented in section 7.

2. Preliminaries

We assume that the true survival times $X_1, \ldots, X_n$ for $n$ items constitute a random sample from an absolutely continuous distribution $F(x) = P[X > x]$ with $F(0) = 1$. The $X_i$ are censored on the right by random variables $C_i$, so that one observes only $Y^*_i = \min(X_i, C_i)$, and an indicator $I(X_i \leq C_i)$, $i=1, \ldots, n$. The $C_i$ are assumed to be a random sample, drawn independently of the $X_i$, from a distribution $G(c) = P[C > c]$. Then the variables $Y^*_1, \ldots, Y^*_n$ form a random sample from the distribution $K = FG$. Let $y_1 \leq \ldots \leq y_n$ denote the ordered observed values of this sample, and set $\delta_i = 1$ if $y_i$ is uncensored, and $\delta_i = 0$ otherwise.

We wish to use the observed $(y_i, \delta_i)$ $i=1, \ldots, n$ to estimate $F$ over an interval $[0, T]$, where $T$ is any finite positive number satisfying $K(T) > 0$. We shall adopt the usual convention of resolving ties between censored and uncensored observations by shifting the censored observations slightly to the left.

The Kaplan-Meier and Nelson-Altshuler estimators $F_{KM}$ and $F_{NA}$ are

$$F_{KM}(t) = \prod_{y_i \leq t} \left( \frac{n-i}{n-i+1} \right)^{\delta_i}.$$
and

\[(2.1) \quad F_{NA}(t) = \exp[-\Lambda_{NA}(t)].\]

Here

\[(2.2) \quad \Lambda_{NA}(t) = \sum_{y_i \leq t} \frac{\delta_i}{n-i+1}\]

estimates the cumulative hazard \(\Lambda(t) = -\ln F(t)\). Kaplan and Meier derived \(F_{KM}\) as the unrestricted maximum likelihood estimate of \(F\). In the absence of tied failures, Breslow (1972) derived \(\Lambda_{NA}\) as the maximum likelihood estimate of \(\Lambda\) in the class of distributions which have constant hazard rates between uncensored observations, with censored observations considered to be censored just after the preceding failure time. In the presence of ties, he obtained the estimator \(F_{Q0}\), discussed in the next section, which differs slightly from \(2.2\). When \(n-i\) is large for all \(y_i \leq t\), the two estimators \(F_{KM}(t)\) and \(F_{NA}(t)\) are in close agreement with each other.

3. The estimators

We shall derive a class of new estimators for \(F\) by the maximum likelihood method. First we note that the likelihood of the data is a product of one factor \(L\) which involves \(F\), and another factor which involves \(G\). Thus to maximize the likelihood with respect to \(F\), we need only maximize the factor \(L\). To do so, we express it as a functional \(L(\lambda)\) of the hazard rate \(\lambda:\)

\[(3.1) \quad L(\lambda) = \prod_{i=1}^{n} [\lambda(y_i)]^{\delta_i} \exp[-\int_{0}^{y_i} \lambda(u)du].\]

Next we approximate \(\lambda(t)\) by a spline function \(\lambda_{Q}(t)\) associated with one of the standard quadrature formulas of numerical integration. For this
purpose we introduce a partition $P$ of the interval $[0,T]$:

$$P: 0 = t_0 < t_1 < \ldots < t_k < t_{k+1} = T.$$  

Then we define $\lambda_Q(t)$ by

$$(3.2) \quad \lambda_Q(t) = \sum_{j=0}^{k+1} \theta_j(t)\lambda_j.$$  

Here $\lambda_j = \lambda(t_j)$ and $\theta_j$ is a polynomial in $t$ in each subinterval $[t_j, t_{j+1}]$, i.e., a spline function. The degree and coefficients of the polynomial vary from one quadrature formula to another, as is shown in the examples below. The integral of $\lambda_Q$, which approximates the cumulative hazard $\Lambda(t)$, is also a spline function of $t$ given by

$$(3.3) \quad \int_0^t \lambda_Q(u)du = \sum_{j=0}^{k+1} \lambda_j \int_0^t \theta_j(u)du = \sum_{j=0}^{k+1} \alpha_j(t)\lambda_j.$$  

The coefficients $\alpha_j(t)$ are splines which do not depend upon $\lambda$, but do depend upon some of the $t_j$.

To use (3.3) we take the subdivision points to be the $k$ distinct failure times $t_j, j=1,\ldots,k$. We denote by $d_j$ the number of failures at time $t_j$. Then in (3.1) we replace the integral of $\lambda$ by that of $\lambda_Q$, use (3.3) and take logarithms to obtain

$$(3.4) \quad \log L(\lambda_Q) = \sum_{j=1}^{k} d_j \log \lambda_j - \sum_{j=0}^{k+1} \gamma_j \lambda_j.$$  

Here

$$(3.5) \quad \gamma_j = \sum_{i=1}^{n} \alpha_j(y_i).$$
We now maximize expression (3.4) with respect to the \( k+2 \) quantities \( \lambda_j \) by equating to zero its \( \lambda_j \) derivatives. The solution of these likelihood equations is

\[
\hat{\lambda}_j = d_j \gamma_j^{-1}, \quad j=1, \ldots, k; \quad \hat{\lambda}_0 = \hat{\lambda}_{k+1} = 0.
\]

When these values \( \hat{\lambda}_j \) are used in (3.2), it becomes the new estimate \( \hat{\lambda}_Q(t) \) of \( \lambda(t) \):

\[
\hat{\lambda}_Q(t) = \sum_{j=0}^{k+1} \theta_j(t) d_j \gamma_j^{-1}.
\]

An estimate for the cumulative hazard is obtained by using (3.6) for \( \lambda_j \) in (3.3):

\[
\hat{\lambda}_0(t) = \sum_{j=1}^{k} \alpha_j(t) d_j \gamma_j^{-1}.
\]

The corresponding estimate for the survivor function is

\[
\hat{F}_Q(t) = \exp\left[- \sum_{j=1}^{k} \alpha_j(t) d_j \gamma_j^{-1}\right].
\]

We shall give some examples of quadrature rules and the corresponding estimators. But first we show that the Nelson-Altshuler estimator (2.1) is a special case of (3.8). To do so we take \( \theta_j(t) = \delta(t-t_j) \), where \( \delta(\cdot) \) is the Dirac delta function. Then \( \lambda_Q \) of (3.2) is

\[
\lambda_Q(t) = \sum_{j=0}^{k+1} \lambda_j \delta(t-t_j),
\]

and \( \alpha_j(t) \) of (3.3) is the Heaviside function

\[
a_j(t) = 0 \quad t < t_j
\]

\[
= 1 \quad t \geq t_j.
\]
Using (3.10) for $\alpha_j$ in (3.5) shows that $\gamma_j$ is the number $N_j$ of observations $y_i$ satisfying $y_i > t_j$. With (3.10) for $\alpha_j$ and $N_j$ for $\gamma_j$, (3.7) and (3.8) become

\begin{equation}
\lambda_{Q_0}(t) = \sum_{t_j \leq t} d_j / N_j ,
\end{equation}

and

\begin{equation}
F_{Q_0}(t) = \exp(- \sum_{t_j \leq t} d_j / N_j).
\end{equation}

Comparison of (3.12) and (2.1) shows that $F_{Q_0}$ agrees with the Nelson-Altshuler estimate $F_{NA}$ in the absence of tied failure times, but differs slightly from it in the presence of ties.

Example 1. Left rectangular rule $Q_1$.

For this rule we take $\theta_0(t) = 0$ and $\theta_j(t)$ to be the indicator function

\begin{equation}
b_j(t) = 1 \quad t_{j-1} < t \leq t_j \\
= 0 \quad \text{otherwise}; \quad j = 1, \ldots, k+1.
\end{equation}

Then $\lambda_Q$ of (3.2) is the left continuous step function

\[ \lambda_{Q_1}(t) = \lambda_{j+1} , \quad t_j < t \leq t_{j+1} , \quad j = 0, \ldots, k \]

shown in Figure 1. By using (3.13) in (3.3), we find that $\alpha_j(t) = a_j(t)$ where $a_j(t)$ is given by $a_0(t) = 0$ and

\begin{equation}
a_j(t) = 0 , \quad 0 \leq t \leq t_{j-1} \\
= t - t_{j-1} , \quad t_{j-1} \leq t \leq t_j \\
= t_j - t_{j-1} = \Delta_j , \quad t_j \leq t \leq T , \quad j = 1, \ldots, k+1.
\end{equation}
Thus $a_j(t)$ is the "time on test" during the interval $[t_{j-1}, t_j]$ of an item observed until $t$. By using $\alpha_j = a_j$ in (3.5), one finds that $\gamma_j$ is the total time on test during the interval $[t_{j-1}, t_j]$ which lies to the left of $t_j$.

Example 2. Right rectangular rule $Q_2$.

For the right rectangular rule, $\theta_j(t)$ is the indicator

\[
\theta_j(t) = 1 \quad t_j \leq t < t_{j+1} \\
= 0 \quad \text{otherwise} \quad j = 0, 1, \ldots, k,
\]

and $\theta_{k+1}(t) = 0$. Then from (3.2) $\lambda_Q$ is the right continuous step function

\[
\lambda_{Q_2}(t) = \lambda_j, \quad t_j \leq t < t_{j+1}, \quad j = 0, \ldots, k
\]

shown in Figure 2. It follows from (3.3) and (3.15) that $\alpha_j(t) = a_{j+1}(t)$, $j=0, \ldots, k$, where $a_{k+2} = 0$. Equation (3.5) shows that $\gamma_j$ is the total time on test during the interval $[t_j, t_{j+1}]$ to the right of $t_j$.

Example 3. Trapezoidal rule $Q_3$.

In this rule, we have

\[
\theta_j(t) = \frac{b_j(t)a_j(t)}{\Delta_j} + \frac{b_{j+1}(t)[\Delta_{j+1} - a_{j+1}(t)]}{\Delta_{j+1}}, \quad j = 1, \ldots, k,
\]

with

\[
\theta_0(t) = \frac{b_1(t)[\Delta_1 - a_1(t)]}{\Delta_1}
\]

and

\[
\theta_{k+1}(t) = \frac{b_{k+1}(t)a_{k+1}(t)}{\Delta_{k+1}}.
\]

Here $b_j(t)$ is the indicator function (3.13), $a_j(t)$ is given by (3.14), and $\Delta_j = t_j - t_{j-1}$. The resulting spline function $\lambda_Q$ of (3.2) is the continuous piecewise linear function.
\[ \lambda_{Q_3}(t) = \frac{t-t_j}{\Delta_{j+1}} \lambda_{j+1} + \frac{t_{j+1}-t}{\Delta_{j+1}} \lambda_j, \quad t_j \leq t \leq t_{j+1}, \quad j = 0, \ldots, k \]

shown in Figure 3. Then (3.3) and (3.16) yield

\[ (3.17) \quad \alpha_j(t) = a_{j+1}(t) + \frac{1}{2}[\Delta_j^{-1}a_j^2(t) - \Delta_{j+1}^{-1}a_{j+1}^2(t)] . \]

When we use (3.17) in (3.5) and (3.6) we obtain for \( \hat{\lambda}_j \) the result

\[ (3.18) \quad \hat{\lambda}_j = d_j/\gamma_j = d_j/[c_{1,j+1} + \frac{1}{2}[\Delta_j^{-1}c_{2j} - \Delta_{j+1}^{-1}c_{2,j+1}]] , \quad j = 1, \ldots, k . \]

In (3.18) we have introduced \( c_{mj} \), the sum of the \( m \)-th powers of the times on test in the interval \( [t_{j-1}, t_j] \), defined by

\[ (3.19) \quad c_{mj} = \sum_{i=1}^{n_m} a_j^m(\gamma_i^m) . \]

In particular \( c_{1j} \) is the total time on test in \( [t_{j-1}, t_j] \) while \( c_{2j} \) is the sum of the squares of the times on test in \( [t_{j-1}, t_j] \).

With this notation, the estimates for the left and right rectangular rules are \( \hat{\lambda}_j = d_j/c_{1j} \) and \( \hat{\lambda}_j = d_j/c_{1,j+1} \) respectively. If the last item in the sample is uncensored, \( c_{1,k+1} = 0 \), and therefore \( \hat{\lambda}_k = \infty \) for the right-rectangular rule. When \( d_j = d_j = 1, \ldots, k \), the estimates \( \Lambda_{Q_1}(t) \) and \( \Lambda_{Q_2}(t) \) are closely related. In fact, from (3.7) we have

\[ (3.20) \quad \Lambda_{Q_2}(t) - \Lambda_{Q_1}(t) = a_{k+1}(t)dc_{1,k+1}^{-1} - a_1(t)dc_{11}^{-1} . \]

Both \( c_{11} \) and \( c_{1,k+1} \), the total time on test in the first and last intervals respectively, increase linearly with the number of items \( n \) almost surely, since we have assumed that \( K(T) > 0 \). Thus the difference (3.20) is \( O_p(n^{-1}) \).

Both rectangular rules are easy to use. By averaging their results for the cumulative hazard, we obtain the new estimators
\( (3.21) \quad \Lambda_{Q_4}(t) = \frac{1}{2}[\Lambda_{Q_1}(t) + \Lambda_{Q_2}(t)] \),

\( (3.22) \quad F_{Q_4}(t) = \exp \left| -\frac{1}{2}[\Lambda_{Q_1}(t) + \Lambda_{Q_2}(t)] \right| . \)

We expect them to be better than the rectangular rule estimators, and this is confirmed in section 5 where their small sample behavior is examined. However the trapezoidal rule estimators are found to be still better, as should be expected.

4. Asymptotic properties of the new estimators

The estimators \( F_Q \) described in the preceding section are uniformly consistent and have the same asymptotic weak convergence properties derived by Breslow and Crowley (1974), Aalen (1976) and Flemington and Harrington (1979) for the Kaplan-Meier and Nelson-Altshuler estimators. To emphasize the sample size \( n \), we let \( \Lambda_{Q,n}(t) \) and \( F_{Q,n}(t) \) denote the estimators \( \Lambda_Q(t) \) and \( F_Q(t) \) introduced in section 3. We assume that failure times are distinct, which is almost surely the case by the assumed continuity of \( F \).

In the appendix we prove the following theorem:

Theorem For \( Q = Q_0, \ldots, Q_4 \)

i) \( \sup_{0 \leq t \leq T} |\Lambda_{Q,n}(t) - \Lambda(t)| \) converges in probability to zero;

ii) the process \( \sqrt{n}[\Lambda_{Q,n}(t) - \Lambda(t)] \) converges weakly to a Gaussian process \( Z_A \) satisfying

\( (4.1) \quad E[Z_A(t)] = 0 \)

\( \text{Cov} [Z_A(s), Z_A(t)] = \int_0^s \frac{dF(u)}{F^2(u)G(u)} , \quad s \leq t . \)
In (4.1), \( \bar{F}(t) = 1 - F(t) \). Properties i) and ii) for \( A_{Q,n} \) lead to analogous properties for \( F_{Q,n} \):

**Corollary** For \( Q = Q_0, \ldots, Q_4 \)

i) \( \sup_{0 \leq t \leq T} |F_{Q,n}(t) - F(t)| \) converges in probability to zero;

ii) the process \( \sqrt{n}[F_{Q,n}(t) - F(t)] \) converges weakly to a Gaussian process \( Z_F \) satisfying

\[
E[Z_F(t)] = 0
\]

\[
\text{Cov}[Z_F(s), Z_F(t)] = F(s)F(t) \int_0^S \frac{d\bar{F}(u)}{F^2(u)G(u)} , \quad s \leq t.
\]

The proof of part i) of the corollary is similar to that of Flemington and Harrington (1979, Theorem). The proof of part ii) can be found in Miller (1981b).

5. Small sample properties

We have used simulated data to examine the small sample behavior of the Kaplan-Meier estimator \( F_{KM} \), and the proposed estimators \( F_Q \) for \( F(t) \). To perform these simulations, we generated 200 sets of observations \( (y_1, \delta_1), \ldots, (y_n, \delta_n) \) for each combination of survival function \( F \), censoring function \( G \) and sample size \( n \) shown in Table 1. There are a total of \( 6 \times 3 \times 4 = 72 \) combinations yielding \( 72 \times 200 = 14,400 \) sets of observations or trials. The six functions \( F \) were chosen to represent hazard rates that are constant, linearly increasing, exponentially increasing, decreasing, and U-shaped and discontinuous. The three functions \( G \) were chosen to represent no censoring, moderate censoring, and heavy censoring.
For each trial we calculated \( F_{KM}(t) \) and \( F_{Q_i}(t) \), \( i = 0,1,\ldots,4 \), at each decile \( t \) of \( F \). The left and right rectangular rule estimators \( F_{Q_1} \) and \( F_{Q_2} \) tended to underestimate and overestimate \( F \), respectively. Therefore we shall not discuss them further. The 200 trials were used to examine the bias, the variance and the root mean square error (RMSE) of the estimates. We illustrate these calculations for the Kaplan-Meier estimator by writing \( \text{av } F_{KM}(t) \) for the mean of the 200 estimates \( F_{KM,i}(t) \) and defining for each decile \( t \)

\[
\text{bias (KM)} = F(t) - \text{av } F_{KM}(t) ,
\]

\[
\text{var (KM)} = \frac{1}{199} \sum_i (F_{KM,i}(t) - \text{av } F_{KM}(t))^2 ,
\]

and \( \text{RMSE (KM)} = [\text{bias}^2(\text{KM}) + \text{var(\text{KM})}]^{\frac{1}{2}} \).

Each of these functions had similar values for all of the five distributions \( F \) shown in Table 1. Thus Table 2 presents selected results only for \( F(t) = \exp(-t^2) \). For untied failure data, when \( F_{NA} = F_{Q_0} \), Flemington and Harrington (1979) found that the RMSE of the Nelson-Altshuler estimator \( F_{NA} \) was slightly smaller than that of \( F_{KM} \) for those \( t \) satisfying \( F(t) \geq 0.20 \), and at times substantially larger when \( F(t) < 0.20 \). We found the same thing. Since \( F_{Q_0} \) never outperformed the estimators \( F_{Q_3} \) and \( F_{Q_4} \) in RMSE, we have not included it in Table 2.

Table 2 illustrates some of the following general findings for small samples \((n \leq 100)\):

i) All estimators are fairly unbiased, except for very small samples \((n = 10)\) with heavy censoring (average censoring proportion \( \pi = .75 \)). The small bias for \( F_{KM} \) agrees with the findings of Chen et al (1982).

ii) The Kaplan-Meier estimator exhibits more bias than do the new estimates at the 10th and 20th percentiles of \( F \), where the data are sparse.

iii) Both new estimators \( F_{Q_3} \) and \( F_{Q_4} \) have smaller variance than do either the Kaplan-Meier or the Nelson-Altshuler estimators, even in the absence of censoring.
iv) The new estimators have smaller RMSE's than do the Kaplan-Meier and Nelson-Altshuler estimators, and the difference increases as the censoring becomes heavier.

v) The trapezoidal estimator $F_{Q_3}$ has smaller RMSE than does the rectangular rule estimator $F_{Q_4}$.

For the survivor function used in Table 2, the new estimators exhibit more bias than does the Kaplan-Meier estimator at the 90th percentile. However the reverse was true for other survivor functions examined.

6. **Estimating survival in the presence of covariates**

Suppose each observation $(y_i, \delta_i, z_i)$ includes a vector $z_i$ of covariates. We assume that these covariates affect survival via a proportional hazards model of the form

\[(6.1) \quad \lambda_*(t; z) = \lambda(t) r(z, \beta).\]

Here $\lambda_*(t; z)$ is the hazard rate for an item with covariates $z_i$, $\beta$ is a vector of unknown parameters, and $r$ is a known nonnegative function satisfying $r(0, \beta) = r(z, 0) = 1$. Thus $\lambda(t)$ is the hazard rate when $z = 0$.

According to (6.1) and (3.1) the likelihood $L(\lambda_*)$ of the data depends both on the baseline hazard rate $\lambda$ and the parameters $\beta$:

\[(6.2) \quad L(\lambda, \beta) = \prod_{i=1}^{n} \{\lambda(y_i) r(z_i, \beta)\}^{\delta_i} \exp\{-r(z_i, \beta)\} \int_0^{y_i} \lambda(u) du \].

Cox (1972) set

\[(6.3) \quad r(z, \beta) = e^{\beta z}\]

and estimated $\beta$ by maximizing a partial log-likelihood

\[(6.4) \quad \sum_{j=1}^{k} [\beta s_j - \log \sum_{\ell \in (R_j, d_j)} \exp(\beta s_\ell)] .\]
Here \( s_j \) is the sum of \( z \) over the \( d_j \) items failing at \( t_j \), and the sum in brackets is taken over all distinct sets of \( d_j \) items from the "risk set" \( R_j \) of those items with \( y_i > t_j \). Cox (1972), Peto (1972) and Efron (1977) have suggested alternatives to (6.4) to resolve computational difficulties that arise in evaluating this sum when the number of ties is not small. These alternatives agree with (6.4) in the absence of ties.

The partial likelihood (6.4) does not depend on \( \lambda \). Cox and others have proposed estimators for \( \lambda(t) \) (or equivalently for \( F(t) = \exp[-\int_0^t \lambda(u)du] \)) using the maximum partial likelihood estimate \( \hat{\beta} \). These estimators typically reduce to the Kaplan-Meier or Nelson-Altshuler estimators when \( \hat{\beta} = 0 \) (see Kalbfleisch and Prentice (1980) for further discussion).

One can use the methods of section 3 to estimate \( \lambda \) and \( \beta \) simultaneously, by maximizing (6.2) with the integral of \( \lambda \) replaced by that of a spline function \( \lambda_Q \). To do so, use (3.3) in (6.2) and take logarithms to obtain

\[
(6.5) \quad \log L = \sum_{j=1}^{k} [d_j \log \lambda_j + \sum_{z=1}^{d_j} \log r(z_j, \beta)] - \sum_{j=0}^{k+1} \Gamma_j(\beta) \lambda_j .
\]

In (6.5) \( \lambda_j = \lambda(t_j) \), where \( t_1, \ldots, t_k \) are the distinct failure times, and \( t_0 = 0, t_{k+1} = T \). The values \( z_j \) index the covariates of the \( d_j \) items failing at \( t_j \), and

\[
(6.6) \quad \Gamma_j(\beta) = \sum_{i=1}^n r(z_i, \beta) \alpha_j(y_i) .
\]

We now maximize \( \log L \) with respect to the \( k+2 \) quantities \( \lambda_j \) and the vector \( \beta \) by equating to zero the corresponding derivatives of the right hand side of (6.5). The solution of the \( \lambda_j \)-derivative equations is

\[
(6.7) \quad \hat{\lambda}_j = d_j/\Gamma_j(\hat{\beta}) j=1, \ldots, k ; \quad \hat{\lambda}_0 = \hat{\lambda}_{k+1} = 0 ,
\]
where \( \hat{\beta} \) is the solution of the \( \beta \)-derivative equations described below.

An estimate for the cumulative hazard is given by using (6.7) for \( \lambda_j \) in (3.3):

\[
(6.8) \quad \Lambda_Q(t) = \sum_{j=1}^{k} \alpha_j(t) d_j / \Gamma_j(\hat{\beta}).
\]

Of course, \( \Lambda_Q(t) \) agrees with the estimator (3.7) when \( \hat{\beta} = 0 \). The likelihood equation for \( \beta \) is

\[
(6.9) \quad \sum_{j=1}^{k} \sum_{i=1}^{d_j} \frac{\nabla_{\beta} r(z_{jL}, \beta)}{r(z_{jL}, \beta)} - \sum_{j=0}^{k+1} \lambda_j \nabla_{\beta} \Gamma_j(\beta) = 0,
\]

where we have used the gradient notation \( \nabla_{\beta} f \) to denote the vector of first partials of \( f \) with respect to the components of \( \beta \). At the maximum point \( \lambda_j = \hat{\lambda}_j \). Therefore we use (6.7) for \( \lambda_j \) in (6.9) to obtain

\[
(6.10) \quad \sum_{j=1}^{k} \left[ \sum_{i=1}^{d_j} \frac{\nabla_{\beta} r(z_{jL}, \beta)}{r(z_{jL}, \beta)} - \frac{d_j \nabla_{\beta} \Gamma_j(\beta)}{\Gamma_j(\beta)} \right] = 0.
\]

This is a vector equation for the estimate \( \hat{\beta} \) of the vector \( \beta \).

When \( r \) is the exponential function (6.3), the equations (6.7) and (6.10) for \( \hat{\lambda}_j \) and \( \hat{\beta} \) become

\[
(6.11) \quad \lambda_j = d_j / \sum_{i=1}^{n} e^{\beta z_i} \alpha_j(y_i), \quad j = 1, \ldots, k
\]

and

\[
(6.12a) \quad \sum_{j=1}^{k} [s_j d_j \sum_{i=1}^{n} w_{ij}(\beta) z_i] = 0.
\]

Here

\[
(6.12b) \quad w_{ij}(\beta) = e^{\beta z_i} \alpha_j(y_i) / \sum_{L=1}^{n} e^{\beta z_L} \alpha_j(y_L).
\]

One can estimate the covariance matrix of \( \hat{\beta} \) by the inverse of the matrix \( I(\beta) \) of negative derivatives of (6.12a), evaluated at \( \hat{\beta} \). The \((\xi, \nu)^{th}\) entry
of $I(\beta)$ is

\begin{equation}
I_{\xi,\nu}(\beta) = \sum_{j=1}^{k} d_j v_{\xi,\nu}^{(j)}, \quad \xi,\nu = 1,\ldots,p,
\end{equation}

where

\begin{equation}
v_{\xi,\nu}^{(j)} = \sum_{i=1}^{n} w_{ij}(\beta) z_i^{(\xi)} z_i^{(\nu)} - \left[ \sum_{i=1}^{n} w_{ij}(\beta) z_i^{(\xi)} \right] \left[ \sum_{i=1}^{n} w_{ij}(\beta) z_i^{(\nu)} \right].
\end{equation}

In (6.13b) $w_{ij}(\beta)$ is given by (6.12b), and $z_i^{(\xi)}$ denotes the $i$th component of the $p$-dimensional vector $z$.

The global null hypothesis $\beta = 0$ can be tested in the usual way via the score statistic $U(0)$, given by the left-hand side of (6.12a) evaluated at $\beta = 0$. We treat $U(0)$ as asymptotically normal with zero mean vector and with covariance matrix $I(0)$. Under the null hypothesis then, the statistic $U^T(0)I(0)^{-1}U(0)$ has an asymptotic chi-squared distribution with $p$ degrees of freedom.

To use (6.11) through (6.13), we must specify the quadrature rule, which determines the $\alpha_j(t)$. Consider first the rule $Q_0$ for which $\lambda_{Q_0}(t)$ is the sum of delta functions and $\alpha_j(t)$ is the Heaviside function (3.10). By using (3.10) in equation (6.11), we find that $\lambda_j$ is the number of failures at $t_j$, divided by the sum over the risk set $R_j$ of the terms $e^{\beta z}$. From (6.8), by using (3.10) for $\alpha_j$ and (6.6) for $T_j$, we obtain the cumulative hazard estimate

\begin{equation}
\Lambda_{Q_0}(t) = \sum_{t_j \leq t} \left( d_j / \sum_{\xi \in R_j} e^{\beta z_i} \right).
\end{equation}

The estimator (6.14) was first derived by Breslow (1972). The equation (6.12) for $\dot{\beta}$ becomes

\begin{equation}
\sum_{j=1}^{k} \left( s_j \cdot d_j \cdot \sum_{\xi \in R_j} z_i^{(\xi)} e^{\beta z_i} / \sum_{\xi \in R_j} e^{\beta z_i} \right) = 0.
\end{equation}
In the absence of ties, (6.15) is the likelihood equation corresponding to Cox's partial likelihood (6.4). In the presence of ties, (6.15) is the likelihood equation for Peto's alternative to (6.4).

Consider next the left rectangular rule \( Q_1 \) for which \( a_j(t) = a_j(t) \) is given by (3.14). Definition (3.14) and equation (6.11) indicate that \( \hat{\lambda}_j \) is the number of failures at \( t_j \), divided by the total "operational time" on test \( \sum_i e^{a_j(y_i)} \) in the interval \([t_{j-1}, t_j] \). The resulting left continuous step function \( \hat{\lambda}_{Q_1} \) is of the same form as the estimate for \( \lambda \) proposed by Oakes (1972). However he estimated \( \beta \) by using Cox's partial likelihood. The weight \( w_{ij} \) in (6.12) is proportional to the operational time contributed by the \( i \)-th item to the interval \([t_{j-1}, t_j] \). In the absence of censoring, (3.14) shows that \( a_j(y_i) = \Delta_j \) for \( y_i \leq t_j \) and \( a_j(y_i) = 0 \) otherwise. Thus (6.12a) reduces to (6.15). In the presence of censoring, however, (6.12a) and (6.15) produce different estimates for \( \beta \).

For the right rectangular rule \( Q_2 \), we have \( a_j(t) = a_{j+1}(t) \). According to (6.11), \( \hat{\lambda}_j \) equals the number of failures at \( t_j \) divided by the operational time on test in \([t_j, t_{j+1}] \). The weights \( w_{ij} \) in (6.12) vanish outside the risk sets \( R_j \), and the weight for an item in \( R_j \) is proportional to its operational time in \([t_j, t_{j+1}] \).

To find the piecewise linear estimate \( \hat{\lambda}_{Q_3}(t) \) obtained by using the trapezoidal rule, we use (3.17) and (6.6) in (6.7) to get

\[
\hat{\lambda}_{Q_3}(t_j) = d_j/[C_{1,j+1} + \frac{1}{2} \Delta_j^{-1} C_{2,j} - \Delta_{j+1}^{-1} C_{2,j+1}], j = 1, \ldots, k,
\]

where

\[
C_{m,j} = C_{m,j}(\beta) = \sum_{i=1}^{n} e^{a_j(y_i)}.
\]
According to (3.17) the weights $w_{ij}$ in (6.12b) vanish for items outside the risk sets $R_{j-1}$. The weight for the $i$th item depends on the first and second powers of the times it spends in the intervals $[t_{j-1}, t_j]$ and $[t_j, t_{j+1}]$:

$$w_{ij}(\beta) = \frac{e^{\beta z_i} [a_{i+1}(y_i) + \frac{1}{2}(\Delta_{j}^{-1} a_{j}(y_i) + \Delta_{j+1}^{-1} a_{j+1}(y_i))]}{C_{1,j+1}^{-1} + \frac{1}{2}(\Delta_{j}^{-1} C_{2,j} - \Delta_{j+1}^{-1} C_{2,j+1})}.$$ 

The results described in the preceding sections suggest that the rectangular and trapezoidal rule estimates for $\beta$ and $\lambda$ will have better small sample properties than those obtained with the delta function rule, and similar large sample properties. They also suggest that the trapezoidal rule estimates will outperform the rectangular rule estimates. The large and small sample properties of these estimates will be described in a separate publication.

7. Example

To illustrate some of the above results, it is convenient to use data of Mays (1972) analyzed by Whittemore and McMillan (1982). Table 3 gives ordered values of time to censoring or failure (death with osteosarcoma) for beagle dogs injected with plutonium. Uncensored values are denoted with asterisks, and doses are given in parentheses. Table 4 outlines the calculations needed to compute the spline function hazard rates $\lambda Q_i(t)$ for $i=0,\ldots,4$ under the null hypothesis that plutonium is unrelated to survival.

Since there are no tied failure times among these data, the hazard rate estimates are $\hat{\lambda}_j = 1/\gamma_j$. The quantities $\gamma_j$ for the left and right rectangular rules $Q_1$ and $Q_2$ are just the total times on test $c_{1,j}$ and $c_{1,j+1}$,
respectively, shown in column 3. The table does not give the total time on test after the last failure, which is \( c_{1,6} = 1531 \) days. The value \( \gamma_j \) for the delta-function rule \( Q_0 \) is simply the number of items on test before the \( j \)th failure, given in column 5. For the trapezoidal rule \( Q_3 \), \( \gamma_j \) is obtained from columns 2-4 using (3.18), and is shown in column 6. The resulting piecewise linear hazard rate \( \lambda_{Q_3}(t) \), plotted in Figure 4, suggests that \( \lambda(t) \) is roughly proportional to a power of \( t \), as is often true of carcinogenesis data.

The survivor function estimators \( F_{KM}, F_{Q3} \) and \( F_{Q4} \) are shown in Figure 5. The spline function estimators \( F_{Q3} \) and \( F_{Q4} \) exceed the Kaplan-Meier (and the Nelson-Altshuler) estimators at each failure time \( t_j \). This difference reflects the former's use of time on test in the intervals \([t_{j-1}, t_j]\) by items censored in the interval. Use of this information decreases the estimated cumulative hazard, and thus increases the survivor function estimate. Thus in samples with heavy censoring, the spline function estimators for \( F \) will tend to exceed the Kaplan-Meier and Nelson-Altshuler estimators.

Table 4 also outlines the calculations needed to compute the score statistic \( U(0) \) and its asymptotic variance, using the delta function rule \( Q_0 \) and the trapezoidal rule \( Q_3 \). Since the likelihood equation (6.15) corresponding to \( Q_0 \) agrees with Cox's partial likelihood equation, the score statistic for \( Q_0 \) is the same as that proposed by Cox. The corresponding chi-squared statistics \( U^2(0)/I(0) \) on one degree of freedom are 4.152 (\( Q_0 \)) and 4.520 (\( Q_3 \)), in close agreement with each other.
8. Discussion

Two features of the proposed spline-function estimators account for their increased efficiency relative to the Kaplan-Meier and Nelson-Altshuler estimators. First, they use the exact censoring times of the censored items. Second, they assume some smoothness properties of the hazard rate function. These features are not unique to the spline function estimators. For uncensored data, Grenander (1956) and Marshall and Proschan (1965) derived maximum likelihood estimators for $F$ in the class of distributions having monotone hazard rate. Padgett and Wei (1980) and Myktyyn and Santner (1981) extended these estimators to censored data and to distributions having a U-shaped hazard rate. They showed strong consistency of the estimators and used simulations to examine their small sample properties.

The increasing hazard rate and right rectangular rule estimators are closely related. Each provides a step function estimate of the hazard rate with jumps at the failure times. Indeed, if the rectangular rule hazard rate estimate is increasing, the two estimates agree. The same close agreement holds for the decreasing hazard rate and left rectangular rule estimators. However the monotone hazard rate estimators are inferior to the rectangular rule estimators in two ways: i) they impose monotonicity constraints on the unknown hazard rate; and ii) they exhibit large bias in small samples even when these constraints are satisfied (Myktyyn and Santner (1981)).

Nonparametric Bayesian estimators for $F$ have been proposed by Susarla and Van Ryzin (1976,1978) and by Ferguson and Phadia (1979). These estimators also use the censoring times and some smoothness properties, like the spline function estimators. However their small sample efficiency has not been investigated.
Anderson and Senthilselvan (1980) have suggested maximizing a penalized log-likelihood function to obtain piecewise smooth estimates of the baseline hazard function in the proportional hazards model (6.1). Their procedure uses the maximum partial likelihood estimate for $\beta$, and it depends on an arbitrary smoothness parameter. The authors do not discuss the small sample properties of the resulting survivor function estimates.

We are now investigating the properties of estimators obtained with spline functions of polynomial degree greater than 1. One could also consider splines whose polynomial degrees are determined by maximum likelihood. The techniques developed by Wahba (1975) may be useful in developing such estimators.

Acknowledgements

The research of A.S.W. was supported by NIH Grants ES00070 and CA23214 and by grants to SIMS from the Sloan Foundation, EPA and NSF, and that of J.B.K. was supported by the National Science Foundation, the Office of Naval Research, the Air Force Office of Scientific Research, and the Army Research Office.
References


References (Cont.)


Appendix

To verify the uniform consistency and the weak convergence properties of the estimators $\Lambda_{Q,n}$, we use the notation of sections 2 and 3. We also introduce the subdistribution functions $\phi$ and $\Gamma$ defined by

$$\phi(t) = P(Y^* \leq t, \delta = 1) = \int_0^t G(u) d\Phi(u)$$

and

$$\Gamma(t) = P(Y^* \leq t, \delta = 0) = \int_0^t F(u) d\tilde{G}(u).$$

It follows that $\phi + \Gamma = \bar{K} = 1-K$. Let $\phi_n, \Gamma_n, \text{ and } \bar{K}_n = 1-K_n$ denote the corresponding empirical (sub)distribution functions, defined by $\phi_n(t) = n^{-1} \sum I(y_i \leq t, \delta_i = 1)$, $\Gamma_n(t) = n^{-1} \sum I(y_i \leq t, \delta_i = 0)$, and $\bar{K}_n(t) = \phi_n(t) + \Gamma_n(t)$. It is convenient to write the Nelson-Altshuler estimator (2.2) as

$$\Lambda_{Q_{0,n}}(t) = \frac{1}{n} \sum_{t_j \leq t} [K_n(t_j)]^{-1}.$$
and to introduce a closely related estimator

\[ \Lambda_n^*(t) = \frac{1}{n} \sum_{j=1}^{[K_n(t-1)]} K_n(t_j-1)_{-1}. \]

For convenience we restate the Theorem, using the standard \( P \) and \( D \) notation for convergence in probability and weak convergence, respectively (Billingsley (1968)).

Theorem: For \( Q = Q_0, \ldots, Q_4 \) the process \( \Lambda_{Q,n} \) satisfies

\[ \sup_{0 \leq t \leq T} |\Lambda_{Q,n}(t) - \Lambda(t)| \overset{P}{\to} 0; \]

\[ \sqrt{n}[\Lambda_{Q,n}(t) - \Lambda(t)] \overset{D}{\to} Z_{\Lambda}(t), \text{ where } Z_{\Lambda} \text{ satisfies (4.1)}. \]

We shall prove the theorem by showing i) that except for terms which are \( o_p(n^{-1}) \), the \( \Lambda_{Q,n} \) are bounded above and below by \( \Lambda_{Q_0,n} \) and \( \Lambda_n^* \), respectively; and ii) that the Theorem holds for \( \Lambda_{Q_0,n} \) and \( \Lambda_n^* \).

Let \( D \) be the space of real valued function on \([0,T]\) having at most jump discontinuities (Billingsley 1968). Under the metric \( \rho \) introduced by Skohorod, \( D \) is a separable metric space. We need not define \( \rho \), but note only that for \( x,y,z \in D \),

\[ \rho(x,y) = \sup_{t \in [0,T]} |x(t) - y(t)|, \]

and

\[ \{x(t) \leq y(t) \leq z(t) \text{ for all } t \in [0,T]\} \Rightarrow \rho(x,y) \leq \rho(x,z) \]

(see Billingsley (1968), p.111). We shall need the following lemmas.

**Lemma 1** The process \( \Lambda_{Q_0,n} \) satisfies (A.3) and (A.4).

**Proof** Aalen (1976; Theorems 6.1, 8.1).
Lemma 2 \[ \sup_{0 \leq t \leq T} |\Lambda_{Q_0,n}(t) - \Lambda_n^*(t)| \] is \( o_p(n^{-k}) \).

Proof It is evident from definitions (A.1) and (A.2) that \( \Lambda_{Q_0,n} - \Lambda_n^* \) is nonnegative and nondecreasing in \( t \). Thus for \( n \geq 1 \), with probability one,

\[
\sup_{0 \leq t \leq T} |\Lambda_{Q_0,n}(t) - \Lambda_n^*(t)| = (\Lambda_{Q_0,n} - \Lambda_n^*)(T).
\]

Let \( \varepsilon_{ij} = 1 - \delta_{ij}, i = 1, \ldots, n, \) let \( i(j) \) be defined so that \( t_j = y_i(j), j = 1, \ldots, k \), and write

\[
\Lambda_{Q_0,n}(T) = \left[ n-i(k+1) \right]^{-1} - \sum_{\ell=1}^{i(k-1)} \frac{(1-\varepsilon_{ij})(n-\ell+1)}{(n-\ell+1)}
\]

and

\[
\Lambda_n^*(T) = n^{-1} - \sum_{\ell=1}^{i(k-1)} \frac{(1-\varepsilon_{ij})(n-\ell)}{(n-\ell)}.
\]

It follows from (A.7) and (A.8) and the identity

\[
\sum_{\ell=1}^{m} \frac{[(n-\ell)(n-\ell+1)]^{-1}}{[(n-\ell)(n-\ell+1)]^{-1}} = m/[n(n-m)],
\]

that

\[
n(\Lambda_{Q_0,n} - \Lambda_n^*)(T) = n \sum_{\ell=1}^{i(k-1)} \frac{\varepsilon_{ij}}{(n-\ell)(n-\ell+1)} + \frac{i(k-1)}{n-i(k+1)} - \frac{i(k-1)}{n-i(k-1)}.
\]

Equation (A.9) can be written in terms of the subdistribution functions \( \phi_n \) and \( \Gamma_n \) as

\[
n(\Lambda_{Q_0,n} - \Lambda_n^*)(T) = \int_{0}^{\phi_n^{-1} \left[ \phi_n(T) - \frac{1}{n} \right]} \frac{1}{K_n(u)K_n(u')} d\Gamma_n(u) + \frac{K_n^{-1} \phi_n(T) - \frac{1}{n}}{K_n \phi_n^{-1} \phi_n(T) + \frac{1}{n}}
\]

\[
- \frac{K_n^{-1} \phi_n(T) - \frac{1}{n}}{K_n \phi_n^{-1} \phi_n(T) - \frac{1}{n}}.
\]

The right-hand side of (A.10) is a continuous functional \( h_n \) of the normalized processes \( V_n = \sqrt{n}[\phi_n - \phi] \) and \( W_n = \sqrt{n}[\Gamma_n - \Gamma] \). That is, \( h_n \) is a continuous
mapping from $D \times D$ into the real line, where $D \times D$ has the product Skorokhod topology. The continuity of $h_{n}$, and the weak convergence properties of $V_{n}$ and $W_{n}$ to Gaussian limit processes $V$ and $W$, imply that

$$(A.11) \quad h_{n}(V_{n},W_{n}) \xrightarrow{p} h(V,W) = \int_{0}^{T} \frac{d\Gamma(u)}{K(u)K(u^{*})}$$

(Billingsley (1968), Theorems 5.5 and 16.4). The assumption $K(T) > 0$ implies that the right hand side of (A.11) is finite. This proves the Lemma.

**Lemma 3** (Billingsley (1968) Theorem 4.1) Let $V_{n}, W_{n}$ and $V$ be random elements of a separable metric space $(S, \rho)$ such that $V_{n}$ and $W_{n}$ have a common domain for each $n$. Suppose that $V_{n}$ converges weakly to $V$, and that $\rho(V_{n}, W_{n})$ converges in probability to zero. Then $W_{n}$ converges weakly to $V$.

**Corollary 1** The process $\Lambda_{n}^{*}$ satisfies (A.3) and (A.4).

**Proof** The relation (A.3) follows immediately from Lemmas 1 and 2, using Slutsky's theorem. Lemmas 1 and 2 combined with (A.5) imply that $V_{n} = n(\Lambda_{Q_{0,n}} - \Lambda)$, $W_{n} = \sqrt{n}(\Lambda_{n}^{*} - \Lambda)$ and $V = Z_{\Lambda}$ satisfy the hypotheses of Lemma 3, which implies (A.4).

**Lemma 4** Let the random variables $t_{ij}c_{mj}$ and $\Delta_{j}$ be defined as in section 2. For $n \geq 1$, $m \geq 1$ and $j=1, \ldots, k+1$, the inequalities

$$(A.12) \quad [nk_{n}(t_{j-1})]^{-1} \leq \Delta_{j}c_{mj}^{-1} \leq [nk_{n}(t_{j})]^{-1}$$

hold with probability 1.

**Proof** The inequalities (A.12) follow by summing over $i=1, \ldots, n$ the inequalities $I(y_{i} \geq t_{j}) \leq a_{j}^{m}(y_{i})\Delta_{j}^{-m} \leq I(y_{i} > t_{j-1})$.

It is convenient to introduce the process

$$(A.13) \quad \tilde{\Lambda}_{n}(t) = \sum_{0 < t_{j} \leq t} \Delta_{j}/c_{lj}$$
It follows from (A.1), (A.2) and (A.12) with \( m = 1 \), that

\[
\Lambda_n^* \leq \Lambda_n \leq \Lambda_{Q_0,n}.
\]

These inequalities, together with lemma 1, Corollary 1, lemma 3 and (A.6), prove

**Corollary 2.** The process \( \Lambda_n \) satisfies (A.3) and (A.4).

Since \( K(T) > 0 \), lemma 4 and the convergence properties of the distribution function \( K_n \) also give

**Corollary 3.** The random variable \( \Delta_j^{-1} \) is \( o_p(n^{-1}) \), \( j = 1, \ldots, k+1, m=1,2, \ldots \).

**Proof of Theorem.** We see from (3.7), (3.14) and (A.13) that

\[
(A.14) \quad |\Lambda_{Q_1,n}(t) - \Lambda_n(t)| = (t-t_{j-1})c_j(-1) \quad t \leq t_k,
\]

\[
0 \quad t > t_k.
\]

Here the random variable \( J = J(t) \) satisfies \( t_{j-1} < t \leq t_j \).

Corollary 3 and (A.14) imply that \( \sup_t \sqrt{n}|\Lambda_{Q_1,n}(t) - \Lambda_n(t)| \xin p 0. \) Thus (A.5),

Corollary 2 and Lemma 3 show that \( \Lambda_{Q_1,n} \) satisfies (A.3) and (A.4).

Equations (3.20) and (3.21) now imply the theorem for the processes \( \Lambda_{Q_2,n} \) and \( \Lambda_{Q_4,n} \) respectively.

Next we define the process

\[
(A.15) \quad \Lambda_n(t) = \frac{1}{2} \sum_{0 < t_j \leq t} \Lambda_j(\Lambda_j + \Lambda_j + 1),
\]
with \( \hat{\lambda}_j \) given by (3.18). Then we use (A.15) and (3.7) with \( \alpha_j \) given by (3.17) and \( \gamma_j \) given by (3.18) to write \( \Lambda_{Q_3,n} = \hat{\lambda} + R_n \), where

\[
2R_n(t) = \Delta_j^{-1} \left[ \hat{\lambda}_j(t-t_j-1)^2 - \hat{\lambda}_{j-1}(t_j-t)^2 \right], \quad t_{j-1} < t \leq t_j.
\]

To verify the theorem for the trapezoidal rule estimator, it suffices to show:

a) \( \sup_t |\tilde{\alpha}_n(t) - \Lambda_{Q_0,n}(t)| = o_p(n^{-\frac{1}{2}}) \);

b) \( \sup_t |R_n(t)| = o_p(n^{-\frac{1}{2}}) \). From lemma 4 and (3.18) we have

\[(A.16) \quad n[p_j K_n(t_j^-) + q_j K_n(t_{j+1}^-)] \leq \frac{2\hat{\lambda}_j^{-1}}{\Delta_j + \Delta_{j+1}} \leq n[p_j K_n(t_{j-1}) + q_j K_n(t_j)] ,\]

where \( p_j = \Delta_j / (\Delta_j + \Delta_{j+1}) \) and \( q_j = 1 - p_j \).

The inequalities (A.16) imply that

\[(A.17) \quad [nK_n(t_{j-1})^{-1}] \leq \hat{\lambda}_j (\Delta_j + \Delta_{j+1}) / 2 \leq [nK_n(t_{j+1}^-)]^{-1} .\]

Summing (A.17) over \( t_j \leq t \) gives

\[(A.18) \quad \Lambda_n(t) \leq \tilde{\alpha}_n(t) \leq \Lambda_{Q_0,n}(t) + E_n(t) ,\]

with

\[(A.19) \quad \sup_t |E_n(t)| \leq [nK_n(T^-)]^{-1} .\]

The inequalities (A.18) and (A.19) and Lemma 2 imply statement (a). To verify statement (b) we use the inequalities

\[ |R_n(t)| \leq \max(\hat{\lambda}_{j-1}, \hat{\lambda}_j) \Delta_j / 2 \]

\[(A.20) \quad \leq \max[\hat{\lambda}_{j-1}(\Delta_{j-1} + \Delta_j) / 2, \hat{\lambda}_j(\Delta_j + \Delta_{j+1}) / 2] .\]

By (A.17) the right-hand side of (A.20) is bounded above by \( [nK_n(t_{j+1}^-)]^{-1} \), which is \( o(n^{-\frac{1}{2}}) \). This completes the proof of the theorem.
## TABLE 1

Survivor and censorship functions \( F \) and \( G \) and sample sizes \( n \) used in simulations

<table>
<thead>
<tr>
<th>( F(t) )</th>
<th>( G(t) )</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e^{-t} )</td>
<td>no censoring</td>
<td>10</td>
</tr>
<tr>
<td>( e^{-t^2} )</td>
<td>( e^{-t} )</td>
<td>25</td>
</tr>
<tr>
<td>( e^{-e^t} )</td>
<td>( e^{-2t} )</td>
<td>50</td>
</tr>
<tr>
<td>( (1+t)^{-1} )</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>( (1+t^2)^{-1} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \begin{cases} e^{-t} &amp; t \leq 1n2 \ 16e^{-5t} &amp; t &gt; 1n2 \end{cases} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>( \pi_G )</td>
<td>F(t)</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>10</td>
<td>.56</td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td>10</td>
<td>.76</td>
<td>.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.10</td>
</tr>
</tbody>
</table>

\[ \pi_G = E_i E_{i1L} \bar{E}_{i1L} / (200n) \], where \( \bar{E}_{i1L} = 1 \) if the \( i^{th} \) observation in the \( z^{th} \) sample is censored and \( \bar{E}_{i1L} = 0 \) otherwise, \( i=1, \ldots, n; z=1, \ldots, 200 \).
<table>
<thead>
<tr>
<th>n</th>
<th>( \pi_G )</th>
<th>F(t)</th>
<th>BIAS</th>
<th>Relative Efficiency</th>
<th>Root Mean Square Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KM</td>
<td>Q_3</td>
<td>Q_4</td>
</tr>
<tr>
<td>25</td>
<td>0</td>
<td></td>
<td>.90</td>
<td>.000</td>
<td>-.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.70</td>
<td>-.005</td>
<td>-.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.50</td>
<td>.003</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.30</td>
<td>.013</td>
<td>.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.10</td>
<td>.003</td>
<td>.001</td>
</tr>
<tr>
<td>25</td>
<td>.55</td>
<td></td>
<td>.90</td>
<td>-.003</td>
<td>-.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.70</td>
<td>.000</td>
<td>.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.50</td>
<td>.009</td>
<td>.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.30</td>
<td>-.002</td>
<td>-.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.10</td>
<td>.019</td>
<td>-.002</td>
</tr>
<tr>
<td>25</td>
<td>.77</td>
<td></td>
<td>.90</td>
<td>.000</td>
<td>-.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.70</td>
<td>-.003</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.50</td>
<td>.021</td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.30</td>
<td>.050</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.10</td>
<td>.127</td>
<td>.042</td>
</tr>
<tr>
<td>n</td>
<td>$\pi_G$</td>
<td>F(t)</td>
<td>BIAS</td>
<td>Relative Efficiency</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>--------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>KM</td>
<td>$V(Q_3)/V(KM)$</td>
<td>$V_{Q_4}/V_{KM}$</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>.90</td>
<td>.001</td>
<td>.001</td>
<td>.938</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.70</td>
<td>.003</td>
<td>.003</td>
<td>.972</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.50</td>
<td>.004</td>
<td>.004</td>
<td>.997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.30</td>
<td>.006</td>
<td>.006</td>
<td>.985</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.10</td>
<td>.003</td>
<td>.003</td>
<td>.950</td>
</tr>
<tr>
<td>100</td>
<td>.55</td>
<td>.90</td>
<td>-.001</td>
<td>-.001</td>
<td>.892</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.70</td>
<td>.003</td>
<td>.003</td>
<td>.980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.50</td>
<td>.003</td>
<td>.003</td>
<td>.941</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.30</td>
<td>-.001</td>
<td>-.001</td>
<td>.943</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.10</td>
<td>.004</td>
<td>-.001</td>
<td>.921</td>
</tr>
<tr>
<td>100</td>
<td>.76</td>
<td>.90</td>
<td>.000</td>
<td>.000</td>
<td>.885</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.70</td>
<td>.004</td>
<td>.003</td>
<td>.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.50</td>
<td>.001</td>
<td>.002</td>
<td>.866</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.30</td>
<td>.004</td>
<td>.001</td>
<td>.850</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.10</td>
<td>.027</td>
<td>.004</td>
<td>.643</td>
</tr>
</tbody>
</table>
TABLE 3

Time in Days to Osteosarcoma (*) or Censoring for 26 Beagles Injected with $^{229}\text{Pu}$. Doses in $\mu\text{Ci} \cdot 10^2/\text{kg}$ are given in Parentheses

1539(1.42), 2061(1.68), 2257*(1.68), 2374(1.67), 3111(1.56), 3111(1.59),
3308*(1.63), 3367*(1.72), 3466(1.51), 3492(1.41), 3495(1.59), 3495(1.63),
3495(1.65), 3539(1.39), 3539(1.40), 3539(1.41), 3649(1.52), 3764(1.68),
3981(1.67), 4292(1.40), 4292*(1.65), 4549(1.39), 4572*(1.50), 4810(1.63),
5160(1.57), 5277(1.53).

<table>
<thead>
<tr>
<th>Failure Time $t_j$</th>
<th>Interval Length $\Delta_j$</th>
<th>Time on Test $c_{m,j}$ $\cdot 10^{-3}$</th>
<th>$c_{1,j} \cdot 10^{-6}$</th>
<th>Quantity $\gamma_j$</th>
<th>Dose $z_j$</th>
<th>$\Sigma_i w_{ij} (0) z_i$ $Q_0$ $Q_3$</th>
<th>$\Sigma_i w_{ij} (0) z_i^2$ $Q_0$ $Q_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2257</td>
<td>2257</td>
<td>57.77</td>
<td>128.87</td>
<td>24</td>
<td>1.68</td>
<td>1.558</td>
<td>1.556</td>
</tr>
<tr>
<td>3308</td>
<td>1051</td>
<td>22.85</td>
<td>23.56</td>
<td>20</td>
<td>1.63</td>
<td>1.544</td>
<td>1.546</td>
</tr>
<tr>
<td>3367</td>
<td>59</td>
<td>1.12</td>
<td>0.07</td>
<td>19</td>
<td>1.72</td>
<td>1.539</td>
<td>1.538</td>
</tr>
<tr>
<td>4292</td>
<td>925</td>
<td>8.88</td>
<td>6.76</td>
<td>6</td>
<td>1.65</td>
<td>1.545</td>
<td>1.533</td>
</tr>
<tr>
<td>4572</td>
<td>280</td>
<td>1.38</td>
<td>0.38</td>
<td>4</td>
<td>1.50</td>
<td>1.558</td>
<td>1.551</td>
</tr>
</tbody>
</table>

$U(0) = \Sigma_j z_j - \Sigma_j \Sigma_i w_{ij} (0) z_i = 0.437$ ($Q_0$); $0.456$ ($Q_3$).

$I(0) = \Sigma_j \Sigma_i w_{ij} (0) z_i^2 - \Sigma_j [\Sigma_i w_{ij} (0) z_i]^2 = 0.046$ ($Q_0$); $0.046$ ($Q_3$).
Figure 1  The piece-wise constant function $\lambda_{Q_1}(t)$ (-----) given by the left rectangular rule, corresponding to the function $\lambda(t)$ (-----).
$\lambda_{Q_1}(t_j) = \lambda(t_j)$, $j=0,1,\ldots$, $k=6,k+1$.

Figure 2  The piece-wise constant function $\lambda_{Q_2}(t)$ (-----) given by the right rectangular rule, corresponding to the function $\lambda(t)$ (-----).
$\lambda_{Q_1}(t_j) = \lambda(t_j)$, $j=0,1,\ldots$, $k=6,k+1$.

Figure 3  The piece-wise linear function $\lambda_{Q_3}(t)$ (-----) given by the trapezoidal rule, corresponding to the function $\lambda(t)$ (-----).
$\lambda_{Q_1}(t_j) = \lambda(t_j)$, $j=0,1,\ldots$, $k=6,k+1$.

Figure 4  Trapezoidal rule estimate $\hat{\lambda}_{Q_3}(t)$ for beagle data.

Figure 5  Kaplan-Meier ($F_{KM}$), trapezoidal rule ($F_{Q_3}$) and rectangular rule ($F_{Q_4}$) estimates of survivor function for beagle data.