▩ 연구논문

Partially Parametric Estimation of Lifetime Distribution from a Record of Failures and Follow-Ups*

Byoung Chang Yoon
Dept. of Mathematics, Korea Military Academy

Abstract

In some observational studies, we have often random censoring model. However, the data available may be partially observable censored data consisting of the observed failure times and only those nonfailure times which are subject to follow-up. In this paper, we present an extension of the problem of partially parametric estimation of the survival function to such partially observable censored data. The proposed estimator treats the observed failure times nonparametrically and uses a parametric model only for those nonfailure times which are subject to follow-up. We discuss the motivation and construction of the proposed estimator and investigate the limiting properties of the proposed estimator such as asymptotic normality. Also, when the assumed parametric model is exponential, the asymptotic variance of the estimator is obtained. Furthermore, an example is given to compare the proposed estimator with the modified Kaplan-Meier (MKM) estimator. From the results, it is shown that the relative efficiency of the proposed estimator is higher than that of the MKM estimator in the follow-up study with increasing time.

1. Introduction

The random censoring model is often used to achieve theoretical results in the area of reliability analysis in engineering and in the area of survival analysis in medical applications. This model is described as follows: Let (X_i, Y_i), $i=1, 2, \cdots, N$, be independent, identically distributed pairs of random variables. X_i is the

* Research supported by the Hwarang-Dae Research Institute, K.M.A.

variable of interest (e.g., the lifetime of an item, the age at death, etc.), and Y_i is some independent censoring variable (e.g., the period of an observational or follow-up study, the warranty period, etc.). The observed quantity is the pair (T_i, δ_i), where

$$T_i = \min(X_i, Y_i)$$
$$\delta_i = I(X_i \le Y_i), i=1, 2, \cdots, N.$$

(Here $I(\cdot)$ denotes the indicator function of the set $\{\cdot\}$). From these pairs (T_i, δ_i), $i=1, 2, \cdots, N$, the lifetime distribution of X_i is estimated.

In some observational studies, not all of the data (T_i, δ_i), $i=1, 2, \cdots, N$, are observed; that is, for some data pairs, only δ_i is observed without the additional information about T_i. Therefore the data set consists of

$$\{(T_i, \delta_i), 1\le i\le n \quad \text{and} \quad \delta_i, n+1\le i\le N\}.$$

This kind of partially observable censored data arises in a variety of situations (cf. Suzuki (1985)).

Suzuki (1985) discussed the problem of nonparametric estimation of the survival function from such partially observable censored data. He compared the modified Kaplan-Meier (MKM) estimator with the Kaplan-Meier (KM; 1958) estimator based only on the follow-up data and the KM estimator applied to all of the data for which T is observed, and showed that the KM estimator applied to the randomly censored data underestimates the survival function and the KM estimator based only on the follow-up data gives a less biased estimate than the KM estimator applied to the randomly censored data but a larger variance than the MKM estimator. Also, he showed that the asymptotic behavior of the estimator is used to study the effect of a follow-up percentage on the precision on the estimator.

Tiwari et al. (1990) discussed the problem of nonparametric Bayes estimation of the survival function for such data of failures and follow-ups under a Dirichlet process prior and squared error loss. They compared the Bayes estimator with the MKM estimator and Susarla and Van Ryzin's (1976) estimator which is based on the randomly censored data (not on the lost observations), and showed that the Bayes estimator is a generalization of the MKM estimator and Susarla and Van Ryzin's estimator underestimates the survival function by using the data which is a modification of the data given in Kaplan and Meier (1958) and used by Susarla and Van Ryzin (1976).

In an observational study of an industrial product, sometimes only the time until failure is observable from the repair requests made by the owner, but nonfailure times are not observable. This paper is concerned with estimation from such partially observable censored data, with making partially parametric assumption.

2. Assumptions and Notations

In this section, we add the assumptions and notations to be used throughout to those in Section 1. Let $\overline{F}(t) \equiv Pr(X_i > t)$ be the survival function of the random variable of interest X_i, and let $\overline{G}(t) \equiv Pr(Y_i > t)$ be the survival function of the censoring random variable Y_i. Assume that X_i and Y_i are independent for all i. Define the survival function of T_i by $S^*(t) = Pr(T_i > t)$, and the subsurvival functions $S_1^*(\cdot)$ and $S_0^*(\cdot)$ by $S_1^*(t) = Pr(T_i > t, \delta_i = 1)$ and $S_0^*(t) = Pr(T_i > t, \delta_i = 0)$, and the subdistribution function $F_0^*(\cdot)$ by $F_0^*(t) = Pr(T_i \le t, \delta_i = 0)$. Note that $S^*(t) = S_1^*(t) + S_0^*(t)$. Furthermore, define the two conditional survival functions of T_i given δ_i by $S_1^c(t) = Pr(T_i > t \mid \delta_i = 1)$ and $S_0^c(t) = Pr(T_i > t \mid \delta_i = 0)$, and the conditional distribution function $F_0^c(\cdot)$ by $F_0^c(t) = Pr(T_i \le t \mid \delta_i = 0)$. Let the known constants $D_i (i = 1, 2, \cdots, N)$ be as follows: $D_i = 1$ if ith item is followed up, and $D_i = 0$ if it is not. If ith item is a failure, then T_i is observed irrespective of the value of D_i, but if it is not a failure then T_i is observed only when $D_i = 1$. Let n_u, n_c, and n_l denote the number of uncensored observations, censored observations that are followed up, and censored observations that are not followed up(lost), respectively. Note that

$$n_u = \sum_{i=1}^{N} \delta_i, \; n_c = \sum_{i=1}^{N} (1-\delta_i) D_i, \; n_l = \sum_{i=1}^{N} (1-\delta_i)(1-D_i).$$

Here $N = n_u + n_c + n_l$.

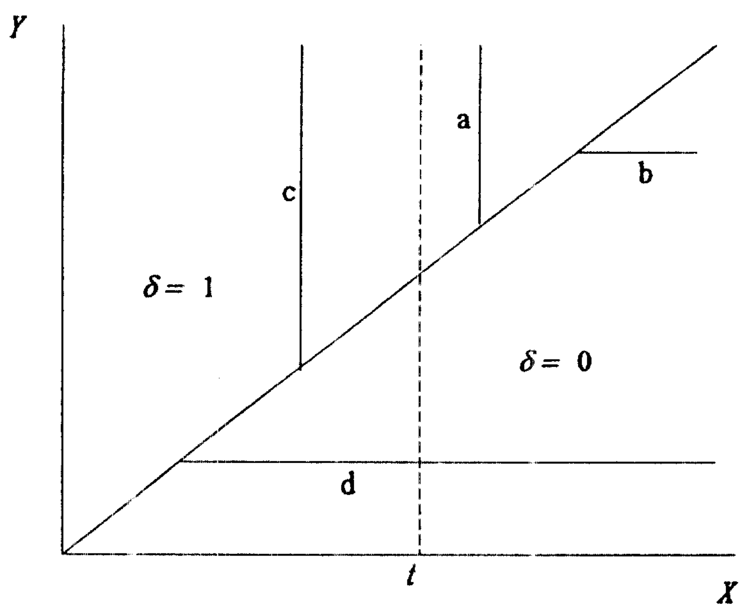
Also, define $n = n_u + n_c$ and $p^* = \sum_{i=1}^{N} D_i / N$, where p^* represents the fraction of items that are followed up. Assume $p^* = n_c / (n_c + n_l)$ and $p^* > 0$.

3. The Proposed Estimator

In this section, we propose a partially parametric estimator of the survival function in the presence of partially observable censored data. Let $S(\cdot \mid \theta)$ be the assumed parametric model for $\overline{F}(\cdot)$, where θ is an unknown parameter. Let $\hat{\theta}$ be a consistent estimator of θ based on the censored sample $\{(T_i, \delta_i), 1 \le i \le n$ and δ_i, $n+1 \le i \le N\}$. The proposed estimator is constructed by analogy to the randomly censored data problem(cf. Klein et al. (1990)). In the partially observable censored data problem we observe points along the line $X=Y=T$ with a ray of possible values of the unobservable coordinate (see Figure 3.1). For those pairs $(T_i, \delta_i=1)$ with $T_i > t$ (ray **a** in Figure 3.1) we are sure that the corresponding X_i is greater than t. For pairs $(T_i, \delta_i=0)$ with $T_i > t$ and D_i (ray **b** in Figure 3.1) the mass of the true X_i being greater than t is $1+n_l/n_c$. For pairs $(T_i, \delta_i=1)$ with $T_i \le t$ (ray **c** in Figure 3.1) we are sure that X_i is less than or equal to t. When $T_i \le t$, $\delta_i=0$ and D_i (ray **d** in Figure 3.1) we cannot determine with certainty if the true unobservable X_i is greater than t or not. In this case an estimate of the chance of the true X_i being greater than t, in light of the observable information (T_i, δ_i) and D_i, is

$$\left(1+\frac{n_l}{n_c}\right)Pr(X_i > t \mid T_i = t_i, \delta_i = 0) = \left(1+\frac{n_l}{n_c}\right)Pr(X_i > t \mid X_i > T_i, Y_i = T_i)$$

which, using the assumed parametric model $S(\cdot \mid \theta)$ for $\overline{F}$, is $(1+n_l/n_c)S(t \mid \hat{\theta})/S(T_i \mid \hat{\theta})$.



⟨ Figure 3.1 ⟩ Four possible rays in the censored data situation.

This suggests the following estimator:

$$\widehat{F}(t) = \sum_{i=1}^{N} \varphi_i(t \mid S(\cdot \mid \hat{\theta}))/N, \tag{3.1}$$

where

$$\varphi_i(t \mid S(\cdot \mid \hat{\theta})) = \begin{cases} 1 & \text{if } T_i > t,\ \delta_i = 1 \\ 1 + \dfrac{n_l}{n_c} & \text{if } T_i > t,\ \delta_i = 0,\ D_i = 1 \\ 0 & \text{if } T_i \le t,\ \delta_i = 1 \\ (1 + \dfrac{n_l}{n_c}) \dfrac{S(t \mid \hat{\theta})}{S(T_i \mid \hat{\theta})} & \text{if } T_i \le t,\ \delta_i = 0,\ D_i = 1. \end{cases} \tag{3.2}$$

The proposed estimator treats the observed failure times nonparametrically and uses a parametric model only for those nonfailure times which are subject to follow-up. Note that Klein's(1990) partially parametric estimator for the randomly censored data is obtained from $\widehat{F}(t)$ by substituting $n_l = 0$. Define $\hat{F}_0^c(t) = (1/n_c) \sum_{i=1}^{N} I(T_i \le t,\ \delta_i = 0) D_i$ and $\hat{F}_0^*(t) = (1/N)(1 + (n_l/n_c)) \sum_{i=1}^{N} I(T_i \le t,\ \delta_i = 0) D_i$. $\hat{F}_0^c(t)$ and $\hat{F}_0^*(t)$ are the generalized maximum likelihood estimators(GMLE's; cf. Kiefer and Wolfowitz(1956)) of $F_0^c(t)$ and $F_0^*(t)$, respectively. Then the proposed estimator is a function of the GMLE's $\hat{S}_1^*(t)$, $\hat{S}_0^*(t)$, and $\hat{F}_0^*(t)$ but it is not the GMLE.

4. Properties of The Estimator

In this section, we investigate the limiting properties of the proposed estimator such as asymptotic normality. Up to the unknown parameter θ, assume that $S(t) = \overline{F}(t)$ for all t. In this case, we can present an alternate expression for $\widehat{F}(t)$ as follows:

$$\widehat{F}(t) = \hat{S}_1^*(t) + \hat{S}_0^*(t) + \int_0^t \frac{S(t \mid \hat{\theta})}{S(u \mid \hat{\theta})} d\hat{F}_0^*(u) \tag{4.1}$$

where

$$\hat{S}_1^*(t) = (1/N) \sum_{i=1}^{N} I(T_i > t,\ \delta_i = 1),$$

$$\hat{S}_0^*(t) = (1/N)(1+(n_l/n_c))\sum_{i=1}^{N} I(T_i>t,\ \delta_i=0)D_i, \text{ and}$$
$$\hat{F}_0^*(t) = (1/N)(1+(n_l/n_c))\sum_{i=1}^{N} I(T_i\le t,\ \delta_i=0)D_i.$$

This representation allows us to prove the following theorem:

Theorem 4.1 *Assume* X_i *and* Y_i *are independent and* $S(t\mid\theta)=\overline{F}(t)$ *under the partially observable censoring model. If* $\hat{\theta}\to\theta$ *with probability* 1, *then* $\hat{\overline{F}}(t)\to\overline{F}(t)$ *uniformly in t with probability* 1.

Proof. First note that $\hat{S}_1^*(t)\to S_1^*(t)=Pr(T_i>t,\ \delta_i=1)=\int_t^\infty[1-G(u)]dF(u)$ uniformly in t with probability 1(w.p.1);
$\hat{S}_0^*(t)=\hat{p}\cdot\hat{S}_0^c(t)\to p\cdot S_0^c(t)=S_0^*(t)=Pr(T_i>t,\ \delta_i=0)=\int_t^\infty[1-F(u)]dG(u)$ uniformly in t w.p.1; $\hat{F}_0^*(t)=\hat{p}\cdot\hat{F}_0^c(t)\to p\cdot F_0^c(t)=F_0^*(t)=Pr(T_i\le t,\ \delta_i=0)$ uniformly in t w.p. 1; and

$$\frac{S(t\mid\hat{\theta})}{S(u\mid\hat{\theta})}\to\frac{\overline{F}(t)}{\overline{F}(u)} \text{ almost surely (a.s.).}$$

Also note that

$$\overline{F}(t) = S_1^*(t)+S_0^*(t)+\int_0^t\frac{\overline{F}(t)}{\overline{F}(u)}dF_0^*(u). \tag{4.2}$$

Let

$$V_{1N}(t) = \hat{S}_1^*(t)-S_1^*(t),\ V_{2N}(t) = \hat{S}_0^*(t)-S_0^*(t),$$
$$V_{3N}(t) = \hat{F}_0^*(t)-F_0^*(t), \tag{4.3}$$

and

$$W_N(t,u) = \frac{S(t\mid\hat{\theta})}{S(u\mid\hat{\theta})}-\frac{\overline{F}(t)}{\overline{F}(u)}. \tag{4.4}$$

It follows that

$$\hat{\overline{F}}(t)-\overline{F}(t) = V_{1N}(t)+V_{2N}(t)+\int_0^t W_N(t,u)dF_0^*(u)$$
$$+\int_0^t\frac{\overline{F}(t)}{\overline{F}(u)}dV_{3N}(u)+\int_0^t W_N(t,u)dV_{3N}(u). \tag{4.5}$$

Since $\hat{\bar{F}}(t)-\bar{F}(t)$ is a continuous function $V_{1N}(t)$, $V_{2N}(t)$, $V_{3N}(t)$, and $W_N(t, u)$ in the supnorm and each of these processes converges to 0 uniformly, the uniform consistency of $\hat{\bar{F}}(t)$ follows. ■

The representation (4.5) allows us to prove, by arguments very closely related to those in Breslow and Crowly(1974), the following weak convergence result.

Theorem 4.2 *Assume X_i and Y_i are independent and $S(t\mid\theta)=\bar{F}(t)$ under the partially observable censoring model. If $\hat{\theta}$ is a consistent estimator of θ and converges in distribution to a normal random variable, then $\sqrt{N}(\hat{\bar{F}}(t)-\bar{F}(t))$ converges weakly to a Gaussian process with mean* 0.

Proof. From (4.5) it follows that

$$\begin{aligned}\sqrt{N}(\hat{\bar{F}}(t)-\bar{F}(t)) &= \sqrt{N}V_{1N}(t)+\sqrt{N}V_{2N}(t)+\int_0^t \sqrt{N}W_N(t,u)dF_0^*(u)\\ &\quad+\sqrt{N}V_{3N}(t)-\bar{F}(t)\int_0^t \sqrt{N}V_{3N}(u)d\bar{F}^{-1}(u)\\ &\quad+\int_0^t \sqrt{N}W_N(t,u)dV_{3N}(u)\\ &= A_N(t)+B_N(t)+C_N(t)+D_N(t)-E_N(t)+R_N(t),\text{ say.}\end{aligned}$$

Expanding $W_N(t, u)$ as a function of $\hat{\theta}$ in a Taylor series about θ yields

$$\sqrt{N}W_N(t,u)=\sqrt{N}(\hat{\theta}-\theta)d[S_0(t\mid\theta)/S_0(u\mid\theta)]/d\theta+o_p(1). \tag{4.7}$$

So $\sqrt{N}W_N(t, u)$ converges to $W(t, u)$, a Gaussian process in t and u. Also, $\sqrt{N}V_{1N}(t)$, $\sqrt{N}V_{2N}(t)$, and $\sqrt{N}V_{3N}(t)$ converge weakly to Gaussian processes, $V_1(t)$, $V_2(t)$, and $V_3(t)$, respectively. Note that $A_N(t)$, $B_N(t)$, $C_N(t)$, $D_N(t)$, and $E_N(t)$ all converge weakly in the supremum metric to Gaussian processes $A(t)$, $B(t)$, $C(t)$, $D(t)$, and $E(t)$, respectively, and $R_N(t)$ converges a.s. to 0 in this metric. By theorem 4.1, $\sqrt{N}(\hat{\bar{F}}(t)-\bar{F}(t))$ converges weakly to a Gaussian process with mean 0 ■

The evaluation of the limiting covariance is difficult, especially for estimators of $\hat{\theta}$ obtained by iterative techniques, since this covariance involves the limiting covariance of

$$\left(\sqrt{N}(\hat{S}_1^*(t)-S_1^*(t)),\sqrt{N}(\hat{\theta}-\theta)\right),\ \left(\sqrt{N}(\hat{S}_0^*(t)-S_0^*(t)),\ \sqrt{N}(\hat{\theta}-\theta)\right),$$
$$\text{and }\left(\sqrt{N}(\hat{F}_0^*(t)-F_0^*(t)),\ \sqrt{N}(\hat{\theta}-\theta)\right).$$

However, this limiting covariance can be obtained in the exponential case given below.

Corollary 4.1 *Assume X_i and Y_i are independent with survival functions $\overline{F}(t)=\exp(-\alpha t)$ and $\overline{G}(t)=\exp(-\beta t)$, respectively, where $\alpha(>0)$ and $\beta(>0)$ are the unknown parameters. If $S(t\mid\hat{\theta})=\exp(-t/\hat{\theta})$ and $\hat{\theta}=\sum_{i=1}^{N}T_i/d$, where d is the observed number of failures ($\hat{\theta}$ is the maximum likelihood estimator (MLE) of $\theta=1/\alpha$), then*

$$Z_N(t)=\sqrt{N}\{\hat{\overline{F}}(t)-\exp(-t/\theta)\}\xrightarrow{L}Z(t)\quad\text{as}\quad N\to\infty$$

(The notation " $\xrightarrow{L}$ " denotes "converges in law to"), *where $Z(t)$ is a Gaussian process with mean 0 and*

$$Var(Z(t))$$

$$=\begin{cases}[\frac{1}{32}+\frac{9}{8}\exp(-2\alpha t)-\frac{7}{4}\alpha t\exp(-2\alpha t)+2\alpha^2t^2 \\ \cdot\exp(-2\alpha t)-2\exp(-3\alpha t)+2\alpha t\exp(-3\alpha t) \\ +\frac{27}{32}\exp(-4\alpha t)+\frac{3}{8}\alpha t\exp(-4\alpha t)] \\ +(\frac{1}{p^*}-1)[-\frac{1}{2}\exp(-2\alpha t)+\alpha t\exp(-2\alpha t) \\ +\exp(-3\alpha t)-\frac{1}{4}\exp(-4\alpha t)] & \text{if}\quad \alpha=\beta \\ [\frac{\beta}{8(\alpha+\beta)^2}+\frac{10\alpha^2-19\alpha\beta+\beta^2}{8\beta(\alpha-\beta)}\exp(-2\alpha t) \\ -\frac{\alpha(2\alpha+9\beta)}{4\beta}t\exp(-2\alpha t)+\alpha(\alpha+\beta)t^2\exp(-2\alpha t) \\ +\frac{\alpha}{\alpha-\beta}\exp\{-(\alpha+\beta)t\}-\frac{2\alpha}{\beta}\exp\{-(2\alpha+\beta)t\} \\ +2\alpha t\exp\{-(2\alpha+\beta)t\}+\frac{\alpha(6\alpha^2+13\alpha\beta+8\beta^2)}{8\beta(\alpha+\beta)^2} \\ \cdot\exp\{-2(\alpha+\beta)t\}+\frac{\alpha^2(2\alpha+\beta)}{4\beta(\alpha+\beta)}t\exp\{-2(\alpha+\beta)t\}] \\ +(\frac{1}{p^*}-1)[-\frac{\alpha}{\alpha-\beta}\exp(-2\alpha t)+\frac{2\beta(\alpha^2+2\alpha\beta-\beta^2)}{(\alpha-\beta)(\alpha+\beta)^2} \\ \cdot\exp\{-(\alpha+\beta)t\}+\frac{2\alpha}{\alpha+\beta}\exp\{-(2\alpha+\beta)t\} \\ -\frac{\alpha^2+2\beta^2}{(\alpha+\beta)^2}\exp\{-2(\alpha+\beta)t\}] & \text{if}\quad \alpha\neq\beta.\end{cases}\tag{4.8}$$

Proof. For the proof of this Corollary, only the evaluation of the limiting variance of $Z(t)$ is needed. First note that $\sqrt{N}(\hat{\theta}-\theta)$ converges in distribution to a normal random variable with mean 0 and variance $(\alpha+\beta)/\alpha^3$. By (4.7) we have

$$\sqrt{N}\,W_N(t, u) = \sqrt{N}(\hat{\theta}-\theta)(t-u)\exp\{-(t-u)/\theta\}/\theta^2 + o(1). \tag{4.9}$$

Hence

$$Cov(W(t, u), W(x, y)) = \alpha(\alpha+\beta)(t-u)(x-y)\exp\{-\alpha(t+x-u-y)\}. \tag{4.10}$$

Now by (4.9)

$$Cov(\sqrt{N}\,V_{1N}(t), \sqrt{N}\,W_N(x, u)) \approx N(x-u)\alpha^2\exp\{-\alpha(x-u)\}Cov(\hat{S}_1^*(t), \hat{\theta}). \tag{4.11}$$

To evaluate $Cov(\hat{S}_1^*(t), \hat{\theta})$ we have

$$\begin{aligned} Cov(\hat{S}_1^*(t), \hat{\theta}) &= E[\hat{S}_1^*(t)\,\hat{\theta}] - E[\hat{S}_1^*(t)]E[\hat{\theta}] \\ &= E[I(T_1>t, \delta_1=1)\,\hat{\theta}] - S_1^*(t)E[\sum_{i=2}^{N} T_i]E[d^{-1}] \\ &= E[\hat{\theta}\mid T_1>t, \delta_1=1]S_1^*(t) - NS_1^*(t)E[d^{-1}]/(\alpha+\beta) \\ &= \{(E[T_1\mid T_1>t, \delta_1=1] + E[\sum_{i=2}^{N} T_i]) - \frac{N}{\alpha+\beta}\}S_1^*(t)E[d^{-1}] \\ &= \{t + \frac{1}{\alpha+\beta} + \frac{N-1}{\alpha+\beta} - \frac{N}{\alpha+\beta}\}S_1^*(t)E[d^{-1}] \\ &= tS_1^*(t)E[d^{-1}]. \end{aligned} \tag{4.12}$$

From Mendenhall and Lehman (1960), we note that

$$E[d^{-1}] \approx (N-2)(\alpha+\beta)/N(N-1)\alpha. \tag{4.13}$$

Combining (4.11), (4.12), and (4.13) and taking the limit as N tends to infinity, we obtain

$$N\,Cov(V_1(t), W(x, u)) = \alpha^2(x-u)t\exp\{-[(\alpha+\beta)t+\alpha(x-u)]\}. \tag{4.14}$$

A similar argument is used to show that

$$N\,Cov(V_2(t), W(x, u)) = \alpha\beta(x-u)t\exp\{-[(\alpha+\beta)t+\alpha(x-u)]\} \tag{4.15}$$

and

$$N\,Cov(V_3(t), W(x, u)) = -\alpha\beta(x-u)t\exp\{-[(\alpha+\beta)t+\alpha(x-u)]\}. \qquad (4.16)$$

Note that, since $p^*=n_c/(n_c+n_t)$, $E[(1+n_t/n_c)\cdot d^{-1}]=(1/p^*)E[d^{-1}]$. By routine calculations, we have the following (for $t<x$):

$$N\,Cov(V_1(t), V_1(x)) = \frac{\alpha}{\alpha+\beta}\Big[\exp\{-(\alpha+\beta)x\} - \frac{\alpha}{\alpha+\beta}\exp\{-(\alpha+\beta)(t+x)\}\Big], \qquad (4.17)$$

$$N\,Cov(V_2(t), V_2(x)) = \frac{1}{p^*}\frac{\beta}{\alpha+\beta}\Big[\exp\{-(\alpha+\beta)x\} - \frac{\beta}{\alpha+\beta}\exp\{-(\alpha+\beta)(t+x)\}\Big], \qquad (4.18)$$

$$N\,Cov(V_3(t), V_3(x)) = \frac{1}{p^*}\frac{\beta}{(\alpha+\beta)^2}\Big[\alpha+\beta\exp\{-(\alpha+\beta)x\} - \alpha\exp\{-(\alpha+\beta)t\} - \beta\exp\{-(\alpha+\beta)(t+x)\}\Big], \qquad (4.19)$$

$$N\,Cov(V_1(t), V_2(x)) = -\frac{\alpha\beta}{(\alpha+\beta)^2}\exp\{-(\alpha+\beta)(t+x)\}, \qquad (4.20)$$

$$N\,Cov(V_1(t), V_3(x)) = \frac{\beta}{(\alpha+\beta)^2}\Big[\exp\{-(\alpha+\beta)t\} - \exp\{-(\alpha+\beta)(t+x)\}\Big], \qquad (4.21)$$

and

$$N\,Cov(V_2(t), V_3(x)) = \frac{1}{p^*}\frac{\beta^2}{(\alpha+\beta)^2}\Big[\exp\{-(\alpha+\beta)t\} - \exp\{-(\alpha+\beta)(t+x)\}\Big]. \qquad (4.22)$$

From the representation (4.6) it follows that

$$\begin{aligned} Var(Z(t)) = {} & Var(A(t))+Var(B(t))+Var(C(t))+Var(D(t))+Var(E(t)) \\ & +2\,Cov(A(t), B(t))+2\,Cov(A(t), C(t))+2\,Cov(A(t), D(t)) \\ & -2\,Cov(A(t), E(t))+2\,Cov(B(t), C(t))+2\,Cov(B(t), D(t)) \\ & -2\,Cov(B(t), E(t))+2\,Cov(C(t), D(t))-2\,Cov(C(t), E(t)) \\ & -2\,Cov(D(t), E(t)). \end{aligned} \qquad (4.23)$$

From (4.17) we have

$$Var(A(t)) = \frac{\alpha}{\alpha+\beta}\left[\exp\{-(\alpha+\beta)t\} - \frac{\alpha}{\alpha+\beta}\exp\{-2(\alpha+\beta)t\}\right]; \tag{4.24}$$

from (4.18),

$$Var(B(t)) = \frac{1}{p^*}\frac{\beta}{\alpha+\beta}\left[\exp\{-(\alpha+\beta)t\} - \frac{\beta}{\alpha+\beta}\exp\{-2(\alpha+\beta)t\}\right]; \tag{4.25}$$

from (4.10),

$$\begin{aligned} Var(C(t)) &= 2\int_0^t \int_0^x Cov(W(t,u), W(t,x))\,dF_0^*(u)\,dF_0^*(x) \\ &= \alpha(\alpha+\beta)\exp(-2\alpha t)\{\exp(-\beta t)+\beta t-1\}^2/\beta^2; \end{aligned} \tag{4.26}$$

from (4.19),

$$\begin{aligned} Var(D(t)) = \frac{1}{p^*}\beta[\alpha+(\beta-\alpha)\exp\{-(\alpha+\beta)t\} \\ -\beta\exp\{-2(\alpha+\beta)t\}]/(\alpha+\beta)^2; \end{aligned} \tag{4.27}$$

from (4.19),

$$Var(E(t)) = \frac{2}{p^*}\int_0^t \int_0^x Cov(V_3(u), V_3(x))\overline{F}(t)^2\,d\overline{F}^{-1}(u)\,d\overline{F}^{-1}(x)$$

$$= \begin{cases} \frac{2}{p^*}\alpha^2\beta\exp(-2\alpha t)\Big[-\frac{(\alpha+\beta)^2}{2\alpha\beta(\alpha-\beta)} + \frac{\exp(2\alpha t)}{2\alpha} \\ \quad + \frac{(\alpha^2+\beta^2)\exp\{(\alpha-\beta)t\}}{\alpha\beta(\alpha-\beta)} - \frac{(\alpha+\beta)\exp(\alpha t)}{\alpha\beta} \\ \quad + \frac{(\alpha+\beta)\exp(-\beta t)}{\alpha\beta} - \frac{\exp(-2\beta t)}{2\beta}\Big]/(\alpha+\beta)^2 & \text{if } \alpha\neq\beta \\ \frac{1}{p^*}\exp(-2\alpha t)\Big[\frac{\exp(2\alpha t)}{4} \\ \quad + \alpha t-\exp(\alpha t)+\exp(-\alpha t)-\frac{\exp(-2\alpha t)}{4}\Big] & \text{if } \alpha=\beta; \end{cases} \tag{4.28}$$

from (4.20),

$$Cov(A(t), B(t)) = -\frac{\alpha\beta}{(\alpha+\beta)^2}\exp\{-2(\alpha+\beta)t\}; \tag{4.29}$$

from (4.14),

$$Cov(A(t), C(t)) = \int_0^t Cov(V_1(t), W(t,u))dF_0^*(u)$$
$$= \alpha^2 t \exp\{-(2\alpha+\beta)t\}[(\beta t-1)+\exp(-\beta t)]/\beta; \quad (4.30)$$

from (4.21),

$$Cov(A(t), D(t)) = \frac{\alpha\beta}{(\alpha+\beta)^2}[-\exp\{-(\alpha+\beta)t\}+\exp\{-2(\alpha+\beta)t\}]; \quad (4.31)$$

from (4.21),

$$Cov(A(t), E(t)) = \int_0^t Cov(V_1(t), V_3(u))\overline{F}(t)d\overline{F}^{-1}(u)$$
$$= -\frac{\alpha\beta}{(\alpha+\beta)^2}\exp\{-(\alpha+\beta)t\}$$
$$+\frac{\alpha}{\alpha+\beta}\exp\{-(2\alpha+\beta)t\} \quad (4.32)$$
$$-\frac{\alpha^2}{(\alpha+\beta)^2}\exp\{-2(\alpha+\beta)t\};$$

from (4.15),

$$Cov(B(t), C(t)) = \int_0^t Cov(V_2(t), W(t,u))dF_0^*(u)$$
$$= \alpha t \exp\{-(2\alpha+\beta)t\}[(\beta t-1)+\exp(-\beta t)]; \quad (4.33)$$

from (4.22),

$$Cov(B(t), D(t)) = \frac{1}{p^*}\frac{\beta^2}{(\alpha+\beta)^2}[-\exp\{-(\alpha+\beta)t\}+\exp\{-2(\alpha+\beta)t\}]; \quad (4.34)$$

from (4.22),

$$Cov(B(t), E(t)) = \frac{1}{p^*}\int_0^t Cov(V_2(t), V_3(u))\overline{F}(t)d\overline{F}^{-1}(u)$$
$$= \frac{1}{p^*}[-\frac{\beta^2}{(\alpha+\beta)^2}\exp\{-(\alpha+\beta)t\}$$
$$+\frac{\beta}{\alpha+\beta}\exp\{-(2\alpha+\beta)t\} \quad (4.35)$$
$$-\frac{\alpha\beta}{(\alpha+\beta)^2}\exp\{-2(\alpha+\beta)t\}];$$

from (4.16)

$$Cov(C(t), D(t)) = \int_0^t Cov(V_3(t), W(t, u))dF_0^*(u)$$
$$= \alpha \exp(-\alpha t)[\beta t - 1 + \exp(-\beta t)][-t\exp\{-(\alpha+\beta)t\}]; \quad (4.36)$$

from (4.16),

$$Cov(C(t), E(t)) = \int_0^t \int_0^x Cov(V_3(u), W(t, x))\bar{F}(t)dF_0^*(u)d\bar{F}^{-1}(x)$$
$$+ \int_0^t \int_0^x Cov(V_3(x), W(t, u))\bar{F}(t)dF_0^*(u)d\bar{F}^{-1}(x)$$
$$= -\frac{\beta^2}{16(\alpha+\beta)^2} + \frac{8\alpha^2 - 2\alpha\beta + \beta^2}{16\beta^2}\exp(-2\alpha t)$$
$$+ \frac{\alpha\beta - 6\alpha^2}{8\beta} t\exp(-2\alpha t) - \frac{\alpha^2}{\beta^2}\exp\{-(2\alpha+\beta)t\} \quad (4.37)$$
$$+ \alpha^2 t^2 \exp\{-(2\alpha+\beta)t\} + \frac{8\alpha^4 + 18\alpha^3\beta + 11\alpha^2\beta^2}{16\beta^2(\alpha+\beta)^2}$$
$$\cdot \exp\{-2(\alpha+\beta)t\} + \frac{6\alpha^3 + 7\alpha^2\beta}{8\beta(\alpha+\beta)} t\exp\{-2(\alpha+\beta)t\};$$

from (4.19),

$$Cov(D(t), E(t)) = \frac{1}{p^*}\int_0^t Cov(V_3(t), V_3(u))\bar{F}(t)\bar{F}^{-1}(u)$$
$$= \frac{1}{p^*}\Big[\frac{\alpha\beta}{(\alpha+\beta)^2} - \frac{\alpha}{\alpha+\beta}\exp(-\alpha t) + \frac{\alpha^2+\beta^2}{(\alpha+\beta)^2}\exp\{-(\alpha+\beta)t\}$$
$$- \frac{\beta}{\alpha+\beta}\exp\{-(2\alpha+\beta)t\} + \frac{\alpha\beta}{(\alpha+\beta)^2}\exp\{-2(\alpha+\beta)t\}\Big]. \quad (4.38)$$

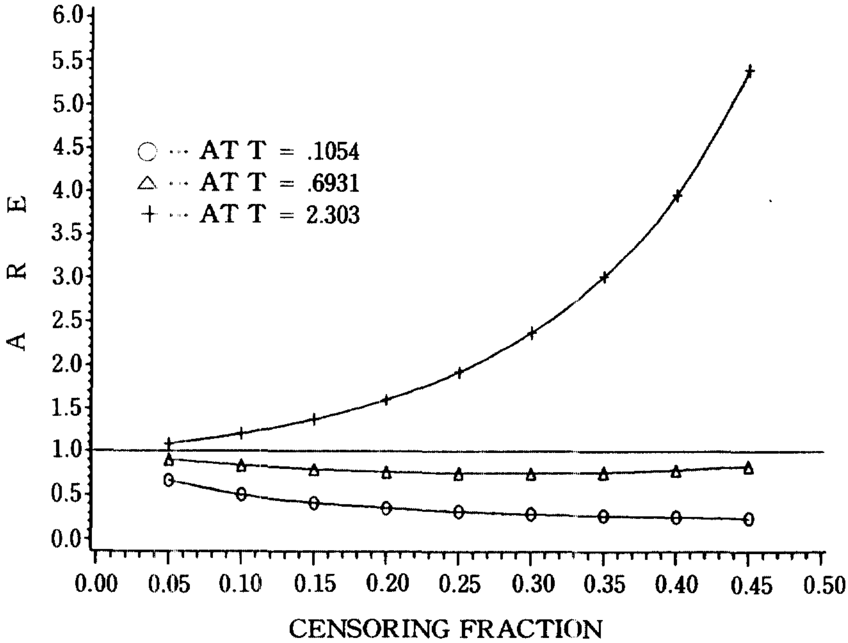
Substituting (4.24) through (4.38) into (4.23), we obtain the result (4.8) after some tedious calculations.

■

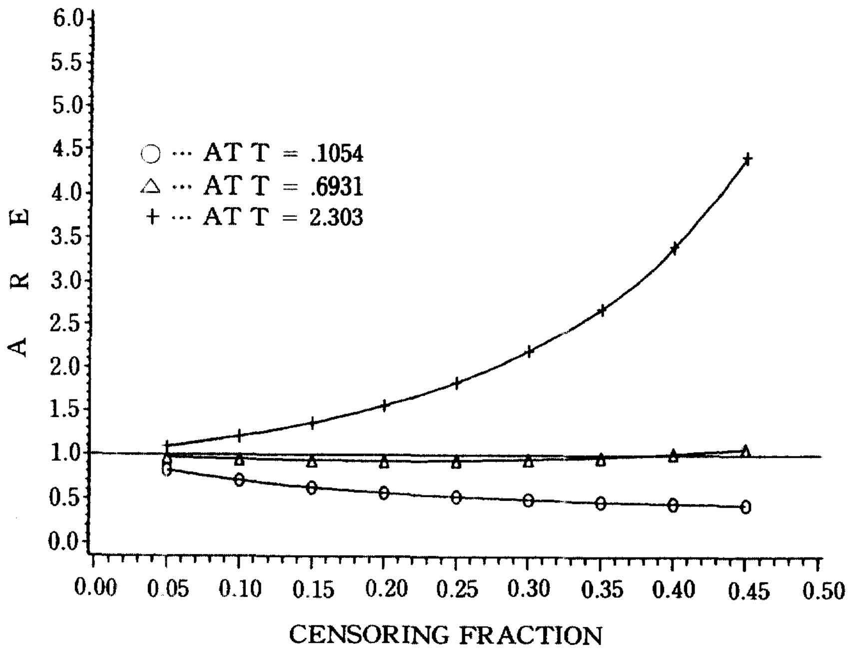
In the exponential case under the condition of partially observable censored data, the partial likelihood is

$$L(\alpha, \beta) = \prod_{i=1}^{N}[\alpha\exp(-\alpha T_i)\cdot\exp(-\beta T_i)]^{\delta_i}[\exp(-\alpha T_i)\cdot\beta\exp(-\beta T_i)]^{1-\delta_i}$$
$$= \alpha^{n_u}\beta^{N-n_u}\exp[-(\alpha+\beta)\sum_{i=1}^{N} T_i].$$

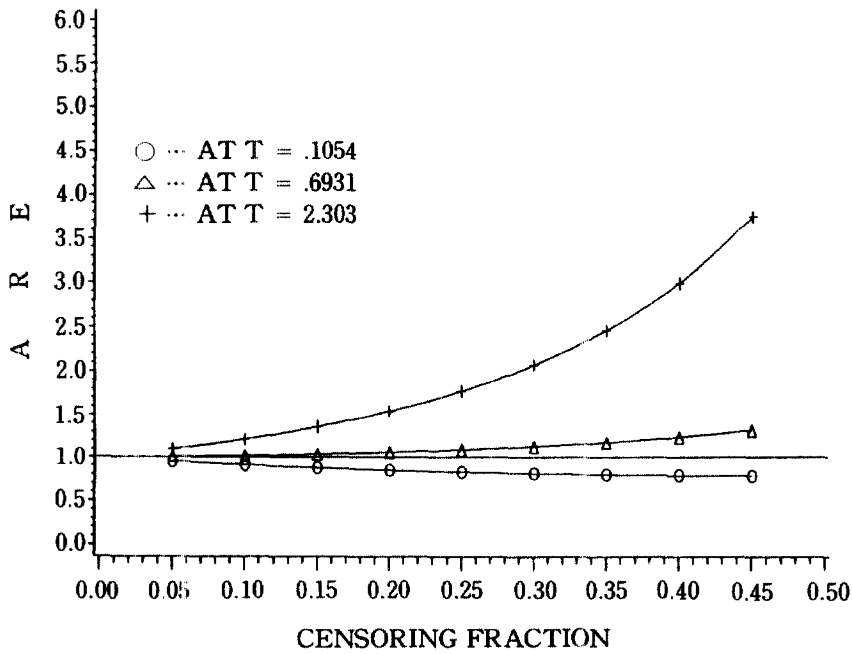
Then, the MLE's of α and β are



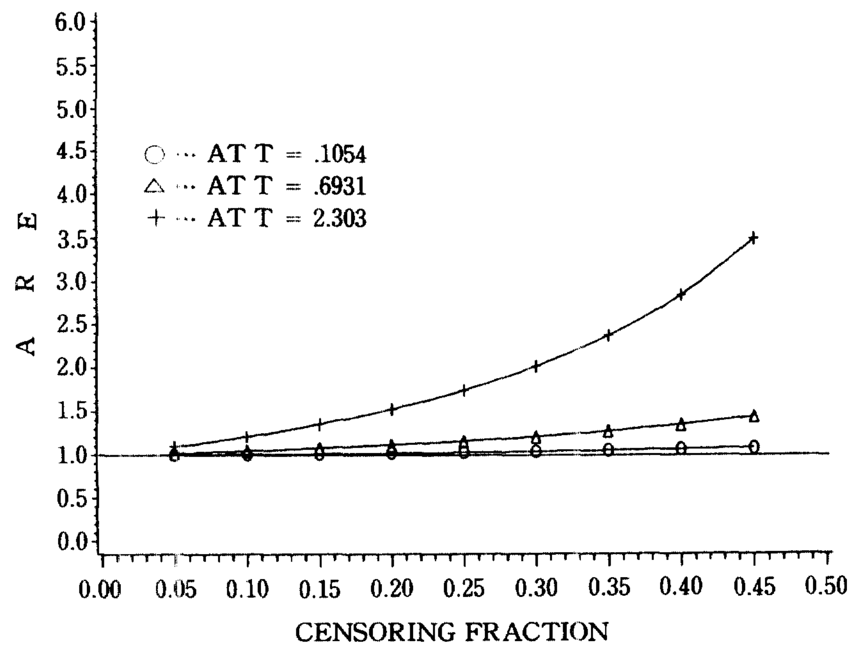
〈 Figure 4.1 〉 Asymptotic relative efficiency as $p^*=0.5$.



〈 Figure 4.2 〉 Asymptotic relative efficiency as $p^*=0.7$.



〈 Figure 4.3 〉 Asymptotic relative efficiency as $p^* = 0.9$.



〈 Figure 4.4 〉 Asymptotic relative efficiency as $p^* = 1.0$.

$\hat{\alpha} = n_u / \sum_{i=1}^{N} T_i$,

$\hat{\beta} = (N - n_u) / \sum_{i=1}^{N} T_i$, respectively.

We can easily obtain a consistent estimator of the asymptotic variance by directly replacing α and β in (4.8) by their MLE's $\hat{\alpha}$ and $\hat{\beta}$.

By taking the limit in (3.3) and (3.4) of Suzuki(1985) as $k \to \infty$ (cf. Tiwari and Zalkikar(1990)), we note that in the exponential case the asymptotic variance of the MKM estimator is

$$\frac{\alpha}{\alpha+\beta} \exp(-2\alpha t) \left[\exp\{(\alpha+\beta)t\} - 1 + \left(\frac{1}{p^*} - 1\right) \cdot \left\{ \frac{\alpha\beta}{\alpha+\beta} t \exp\{(\alpha+\beta)t\} - \alpha\beta t^2 - \frac{\alpha\beta}{\alpha+\beta} t \right\} \right], \tag{4.39}$$

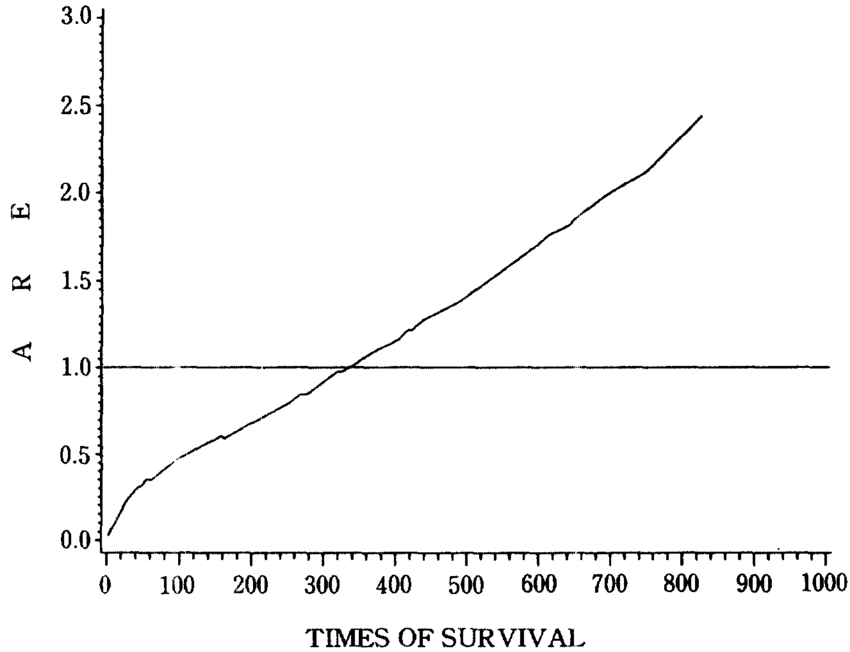
which is always greater than (4.8) when $p^* = 1$. Figures 4.1, 4.2, 4.3, and 4.4 show plots of the asymptotic relative efficiency(ARE) for various values of p^* as a function of the censoring fraction $p = \beta / (\alpha + \beta)$ for $\alpha = 1$, $0 \le p \le .5$, at $t = .1054$, $.6931$, and 2.303, that is, 10^{th}, 50^{th}, and 90^{th} percentiles of the survival distribution, where the ARE is (4.39)/(4.8). From these figures, note that the relative efficiency of our proposed estimator improves with increased censoring subject to follow-up and increasing time. From the comparison of the effect of the fraction of follow-up and the percentile of the survival distribution, it is shown that the relative efficiency of the proposed estimator is always higher than that of the MKM estimator at 90^{th} percentile of the survival distribution regardless of the value of p^*. However, the relative efficiency of the proposed estimator tends to decrease as p^* increases. For small percentile of the survival distribution, the relative efficiency of the proposed estimator is lower than that of the MKM estimator. But the proposed estimator performs better as p^* increases.

5. An Example

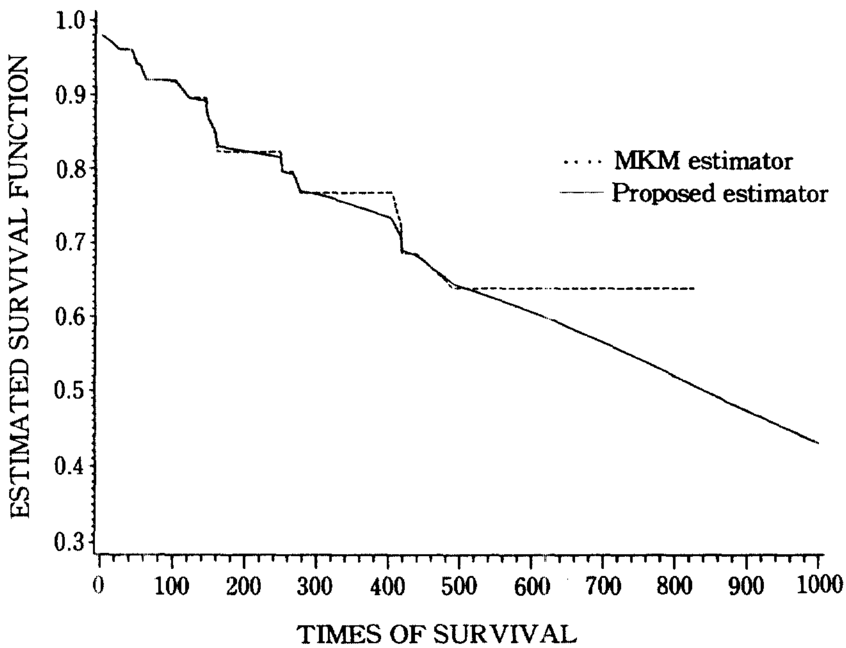
In this section we compare the proposed estimator $\hat{\hat{F}}$ with the MKM estimator $\hat{\hat{F}}$ in the exponential case. This comparison is done by using the data in ⟨Table 5.1⟩ which is a modification of the data given in Klein et al.(1990). Note that the total number of observations is 51, of which 42 are observed T's with 13 failures and 29 observed nonfailures(i.e. $N=51$, $n_u=13$, $n_c=29$, $n_l=9$). In this study 39 items were followed up and hence $p^* = 39/51 \doteqdot 0.76$. From the data in ⟨Table 5.1⟩, we obtain $\hat{\alpha} = 1/\hat{\theta} = 9.38 \times 10^{-4}$, $\hat{\beta} = 2.74 \times 10^{-3}$ which are the MLE's of α and β in

〈 Table 5.1 〉 Failures and Follow-up data:
The proposed estimates $\hat{F}(\cdot)$, the MKM estimates $\hat{F}_s(\cdot)$,
$\sqrt{avar\,\hat{F}(\cdot)}$, $\sqrt{avar\,\hat{F}_s(\cdot)}$, and the ARE

N	T_i	δ_i	D_i	$\hat{F}(\cdot)$	$\hat{F}_s(\cdot)$	$\sqrt{avar\,\hat{F}(\cdot)}$	$\sqrt{avar\,\hat{F}_s(\cdot)}$	ARE
1	2	1	1	.980	.980	.037	.006	.026
2	27	1	1	.961	.961	.047	.022	.219
3	32	0	1	.961	.961	.048	.024	.250
4	43	0	1	.961	.961	.051	.028	.301
5	50	1	0	.941	.940	.053	.030	.320
6	55	0	1	.940	.940	.054	.032	.351
7	62	1	1	.920	.919	.056	.033	.347
8	82	0	1	.919	.919	.059	.038	.415
9	102	0	1	.917	.919	.062	.043	.481
10	103	0	1	.917	.919	.062	.043	.481
11	122	1	1	.895	.895	.065	.047	.523
12	145	0	1	.891	.895	.067	.051	.579
13	148	1	1	.871	.871	.067	.051	.579
14	158	1	0	.850	.847	.068	.053	.607
15	162	1	1	.830	.822	.069	.053	.590
16	194	0	1	.825	.822	.071	.058	.667
17	250	0	1	.815	.822	.074	.066	.795
18	251	1	1	.796	.796	.074	.066	.795
19	267	0	1	.793	.796	.074	.068	.844
20	276	1	1	.771	.768	.075	.069	.846
21	284	0	1	.769	.768	.075	.070	.871
22	292	0	1	.768	.768	.075	.071	.896
23	319	0	1	.761	.768	.076	.075	.974
24	326	0	1	.759	.768	.076	.075	.974
25	346	0	1	.753	.768	.077	.078	1.026
26	365	0	1	.747	.768	.077	.080	1.079
27	404	0	1	.734	.768	.078	.084	1.160
28	417	1	1	.710	.727	.078	.086	1.216
29	418	1	0	.690	.686	.078	.086	1.216
30	423	0	1	.689	.686	.078	.086	1.216
31	438	0	1	.683	.686	.078	.088	1.273
32	491	1	1	.644	.638	.079	.093	1.386
33	584	0	1	.612	.638	.080	.103	1.658
34	595	0	1	.608	.638	.080	.104	1.690
35	613	0	1	.601	.638	.080	.106	1.756
36	642	0	1	.590	.638	.081	.109	1.811
37	649	0	1	.587	.638	.081	.110	1.844
38	693	0	1	.568	.638	.081	.114	1.981
39	707	0	1	.562	.638	.081	.115	2.016
40	746	0	1	.545	.638	.082	.119	2.106
41	755	0	1	.541	.638	.082	.120	2.142
42	826	0	1	.508	.638	.082	.128	2.467



〈 Figure 5.1 〉 Asymptotic relative efficiency.



〈 Figure 5.2 〉 Estimated survival function for MKM estimates and proposed estimates.

Corollary 4.1. In ⟨Table 5.1⟩, the proposed estimates $\hat{F}(\cdot)$ and the MKM estimates $\hat{F}_s(\cdot)$ are given. The approximations of $\sqrt{avar\,\hat{F}(\cdot)}$ and $\sqrt{avar\,\hat{F}_s(\cdot)}$ and the ARE are also given. Here, $avar[\cdot]$ denotes the asymptotic variance of the estimator.

⟨Figure 5.1⟩ displays the ratio of the approximation of $avar\ \hat{F}_s(\cdot)$ to the approximation of $avar\ \hat{F}(\cdot)$ in ⟨Table 5.1⟩ at the time of the survival function. From this figure, we note that the relative efficiency of $\hat{F}(\cdot)$ increases as the time increases. The proposed estimates $\hat{F}(\cdot)$ and the MKM estimates $\hat{F}_s(\cdot)$ are plotted in ⟨Figure 5.2⟩. Note that $\hat{F}(\cdot)$ is defined for all t where as $\hat{F}_s(\cdot)$ is undefined in the right tail when the largest observation is censored. In the plot in the figure, $\hat{F}(\cdot)$ is seen to be smoother than $\hat{F}_s(\cdot)$ in that the jumps at the uncensored observations are not as large for $\hat{F}(\cdot)$.

6. Conclusion and Remarks

In this paper, we proposed a partially parametric estimator of the survival function in the presence of partially observable censored data. The partially parametric estimator discussed by Klein et al. retained most of the distribution-free properties of the KM estimator and, yet, allowed one to estimate the function with reasonable accuracy in the tails. The partially observable censored data discussed by Suzuki would arise in a variety of situations. Thus, the proposed estimator treated the observed failure times nonparametrically and used a parametric model only for those nonfailure times which are subject to follow-up.

In this paper, the asymptotic distribution of the proposed estimator was obtained under the assumption that the assumed parametric model is exponential. Also, the asymptotic variance, which is always less than that of the MKM estimator in the exponential case when $p^*=1$, was obtained after some tedious calculations. The ARE of our proposed estimator improved with increased censoring subject to follow-up and increasing time but it did not improve with increased follow-up, especially at large t.

In conclusion, our proposed estimator based on the exponential distribution for $S(\cdot\mid\theta)$ performs better for the follow-up study with increasing time. But, our estimator performs more poorly for small percent follow-up at small percentile of survival function. Also, the evaluation of the asymptotic variance of the proposed estimator under general parametric model is difficult. This problem will be remained for further studies. Many case studies for partially observable censored data also will be remained.

References

[1] Bain, L. J. (1978), *Statistical Analysis of Reliability and Life Test Models*, New York: Marcel Dekker.

[2] Breslow, N. and Crowley, J. (1974), "A large sample study of the life table and product limit estimates under random censorship," *Annals of Statistics* 2, pp. 437–453.

[3] Grab, E. L. and Savage, J. R. (1954), "Tables of the expected value of $1/X$ for positive Bernoulli and Poisson variables," *Journal of the American Statistical Association* 58, pp. 468–473.

[4] Kaplan, E. and Meier, P. (1958), "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association* 53, pp. 457–481.

[5] Kiefer, J. and Wolfowitz, J. (1956), "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters," *Annals of Mathematical Statistics* 27, pp. 887–906.

[6] Klein, J. P., Lee, S., and Moeschberger, M. L. (1990), "A partially parametric estimator of survival in the presence of randomly censored data," *Biometrics* 46, pp. 795–811.

[7] Miller, R. G., Jr. (1981), *Survival analysis*, New York: Wiley.

[8] Mendenhall, W. and Lehman, E. H., Jr. (1960), "An approximation to the negative moments of the positive binomial useful in life testing," *Technometrics* 2, pp. 227–242.

[9] Peterson, A. V. (1977), "Expressing the Kaplan-Meier estimator as a function of empirical survival functions," *Journal of the American Statistical Association* 72, pp. 854–858.

[10] Susarla, V. and Van Ryzin, J. (1976), "Nonparametric Bayesian estimation of survival curves from incomplete observations," *Journal of the American Statistical Association* 61, pp. 897–902.

[11] Suzuki, K. (1985), "Nonparametric estimation of lifetime distributions from a record of failures and follow-ups," *Journal of the American Statistical Association* 80, pp. 68–72.

[12] Tiwari, R. C. and Zalkikar, J. N. (1990), "Nonparametric Bayes estimation of the survival function from a record of failures and follow-ups," *Communications in Statistics-Theory and Methods* 19(11), pp. 4331–4354.