ESTIMATION OF SURVIVAL FUNCTION BASED ON MODELING OF CENSORING PATTERN

Akio Suzukawa* and Nobuhiro Taneichi**

The Kaplan-Meier estimator (KM-estimator) is an important tool in the analysis of right censored data. It is a non-parametric estimator of an unknown survival function of a lifetime random variable. The purpose of this paper is to obtain a semi-parametric estimator of the survival function. In many practical data, there are several patterns of censoring, for example, censoring is apt to occur for the larger observable time. Such a pattern can be expressed by a function defined by conditional probability of censoring under condition that the observable time is given. We call the function censoring pattern function, and assume a parametric form for it. The survival function estimator derived in this paper is semi-parametric in the sense that the parametric form is assumed only for the censoring pattern function. The estimator is a generalization of the KM-estimator, and the estimated variance derived in this paper is a generalization of the Greenwood's formula. We investigate asymptotic properties of the semi-parametric estimator and show that it is asymptotically more efficient than the KM-estimator. We also show that its mean absolute error is smaller than that of the KM-estimator by a simulation study.

Key Words and Phrases: censored data problem, Kaplan-Meier estimator, logistic regression, survival data.

1. Introduction

Over the past decades a considerable number of studies have been made on analyses of the censored data. Particular attention has been given to developing nonparametric methods for the right censored data. The Kaplan-Meier estimator (KM-estimator) developed by Kaplan & Meier (1958) is the most fundamental tool, and it gives the nonparametric maximum likelihood estimator of an unknown survival function as shown by Kaplan & Meier (1958) and Johansen (1978). Several authors have investigated properties of it (Efron, 1967; Breslow & Crowley, 1974; Gill, 1980, 1983; Reid, 1981; Wang, 1987).

This paper is concerned with the randomly right censored data problem. For $i=1, 2, \dots, n$, let X_i and Y_i be nonnegative random variables which represent a lifetime and a censoring time of the i th individual, respectively. We assume that (X_i, Y_i) , $i=1, 2, \dots, n$ are independent identically distributed pairs of random variables, and X_i 's and Y_i 's are independent. Denote by $F(x)=\text{pr}(X_i \leq x)$ and $G(y)=\text{pr}(Y_i \leq y)$ the distribution functions of the random variables X_i and Y_i , respectively, and let $\bar{F}(x)=1-F(x)$ and $\bar{G}(y)=1-G(y)$ be the

Received November, 1999. Revised May, 2000. Accepted May, 2000.

Department of Animal Production and Agricultural Economics, Obihiro University, Inada-cho, Obihiro, 080-8555, Japan.

* E-mail: suzukawa@obihiro.ac.jp

** E-mail: nobutane@obihiro.ac.jp

survival functions of X_i and Y_i , respectively. In the randomly right censored data, the pairs (X_i, Y_i) , $i=1, 2, \dots, n$ are not observed, and pairs of random variables

$$(Z_i, \delta_i) = (\min(X_i, Y_i), I(X_i \leq Y_i)), \quad i=1, 2, \dots, n$$

are observable, where $I(A)$ denotes the indicator function of a set A .

As Miller (1983) pointed out, the KM-estimator can be inefficient compared to parametric estimator. In parametric analysis of the censored data, a parametric form is assumed only for F in which we are interested. A good point of this parametric approach is that we can make a likelihood-inference without any assumption for G . A weak point of this approach is that checking adequacy for the parametric form assumed for F is difficult, especially for heavily censored data. Model selection criteria such as AIC (Akaike, 1973) and BIC (Schwarz, 1978) can not be directly applied to this parametric model.

Table 1 (Freireich et al., 1963; Gehan, 1965) shows times of remission of 21 leukemia patients with the drug 6-mercaptopurine (6-MP). In this data, we can perceive a pattern of censoring suggesting that censoring is apt to occur for the larger observed time. In other words, the conditional probability of $\delta=0$ (censoring) for given $Z=t$ is near to one for a large t . Such a pattern of censoring can be described by a function

$$m(t) = E(\delta_i | Z_i = t) = \text{pr}(\delta_i = 1 | Z_i = t).$$

We call $m(t)$ censoring pattern function, and assume a parametric form for $m(t)$. Denote by $H(t) = 1 - \bar{F}(t)\bar{G}(t)$ and $\bar{H}(t) = \bar{F}(t)\bar{G}(t)$ the distribution function and the survival function of Z_i , respectively. Then the distribution of the observable pair (Z_i, δ_i) is completely determined by m and H . Hence, the model in which the parametric form with finite dimensional parameters is assumed only for m can be considered as a semi-parametric model. For this semi-parametric model, we can directly use the model selection criteria by considering that the Z_i 's are fixed. We also note that the function m plays a crucial role in Stute & Wang (1993) and Stute (1994) which discussed properties of a Kaplan-Meier integral.

The most fundamental model for the function $m(t)$ is a model in which the censoring pattern is assumed to be constant, i.e., $m(t) = q$ ($0 < q < 1$). If $\bar{H}(t)$ is continuous, then

$$(1.1) \quad \bar{F}(t) = \{\bar{H}(t)\}^q \quad \text{and} \quad \bar{G}(t) = \{\bar{H}(t)\}^{1-q},$$

and hence $\{\bar{G}(t)\}^q = \{\bar{F}(t)\}^{1-q}$, which is called the Koziol-Green proportional hazard model. If $\bar{H}(t)$ is discrete with atoms $\{h_j\}$ at points $\{a_j\}$, $a_1 < a_2 < \dots$,

$$(1.2) \quad \bar{F}(t) = \prod_{j: a_j \leq t} (1 - q\lambda_j) \quad \text{and} \quad \bar{G}(t) = \prod_{j: a_j \leq t} \{1 - (1-q)(1 - q\lambda_j)^{-1}\lambda_j\},$$

where $\lambda_j = h_j/\bar{H}(a_j-)$. Under the assumption $m(t) = q$, a natural estimator of the parameter q is $\hat{q} = \sum_{i=1}^n \delta_i/n$, which is the maximum likelihood estimator. When $\bar{H}(t)$ is estimated by a suitable continuous survival function, then (1.1)

gives estimators of $\bar{F}$ and $\bar{G}$ by substituting $\hat{q}$ into q . On the other hand, if $\bar{H}(t)$ is estimated by a discrete survival function, estimators of $\bar{F}$ and $\bar{G}$ are given by (1.2).

Let $z_{(1)} < z_{(2)} < \cdots < z_{(k)}$ be k distinct values of observed Z_i 's, and there are d_{1j} lifetimes and d_{0j} censoring times at $z_{(j)}$. Put $d_j = d_{1j} + d_{0j}$ and denote by n_j the number of individuals at risk at $z_{(j)}$. When no information about $\bar{H}$ is available, $\bar{H}(t)$ is estimated by an empirical survival function

$$(1.3) \quad \bar{H}_n(t) = 1 - H_n(t) = \prod_{j: z_{(j)} \leq t} (1 - d_j/n_j),$$

which is a discrete survival function with hazard $\hat{\lambda}_j = d_j/n_j$ at $z_{(j)}$. Hence, under the assumption $m(t) = q$, the survival function $\bar{F}(t)$ may be estimated by

$$(1.4) \quad \prod_{j: z_{(j)} \leq t} (1 - \hat{q} d_j/n_j).$$

If the assumption $m(t) = q$ is true, this estimator may have better performance than the KM-estimator

$$(1.5) \quad \bar{F}_n(t) = 1 - F_n(t) = \prod_{j: z_{(j)} \leq t} (1 - d_{1j}/n_j).$$

This drives us to consider estimating the survival function based on modeling of the censoring pattern.

In §2, we derive a semi-parametric estimator of the survival function $\bar{F}$ based on parametric modeling of the censoring pattern function $m(t)$, and derive estimated variance of it. In §3, we investigate asymptotic properties of the semi-parametric estimator and discuss large sample comparison with the KM-estimator. In §4, we illustrate how the semi-parametric estimator can be applied to the leukemia data of Table 1. In §5, a simulation study of the semi-parametric estimator and the KM-estimator is presented.

Table 1. Times of remission (weeks) of leukemia patients with 6-MP

6	6	6	6 ⁺	7	9 ⁺	10	10 ⁺	11 ⁺	13
16	17 ⁺	19 ⁺	20 ⁺	22	23	25 ⁺	32 ⁺	32 ⁺	34 ⁺
35 ⁺									

⁺ Censored observation

2. Survival and censoring pattern functions

2.1. Relation between (F, G) and (m, H) when H is discrete

We shall now look carefully into the relation between (F, G) and (m, H) when H is discrete distribution. Assume that H is discrete with jumps $h_j > 0$ at $a_j, j=1, \dots, k$, where $a_1 < \cdots < a_k$. Then the censoring pattern function $m(t)$ is defined at $t = a_j, j=1, \dots, k$. From the relation $\bar{H}(t) = \bar{F}(t)\bar{G}(t)$, we can see that, for $0 \leq t \leq a_k$, $\bar{F}(t)$ and $\bar{G}(t)$ are right continuous step functions which have no jumps at points except $a_1, \dots, a_k$. Therefore, for $0 \leq t \leq a_k$, $\bar{F}(t)$ and $\bar{G}(t)$ can be expressed as

$$(2.1) \quad \bar{F}(t) = \prod_{j: a_j \leq t} (1 - \lambda_j^F) \quad \text{and} \quad \bar{G}(t) = \prod_{j: a_j \leq t} (1 - \lambda_j^G),$$

where λ_j^F and λ_j^G are hazards of F and G at a_j , respectively. Moreover we shall notice that either of $\bar{F}(a_k)$ and $\bar{G}(a_k)$ must be zero. This implies that either of λ_k^F and λ_k^G is one.

When $\lambda_j^F, \lambda_j^G, j=1, \dots, k$ are given and $\lambda_k^F=1$, from (2.1), we can know about $\bar{F}(t)$ for all t , however $\bar{G}(t)$ is not determined for $t > a_k$. Conversely, when $\lambda_k^G=1$, $\bar{F}(t)$ is not determined for $t > a_k$. In short, for $t > a_k$, the one is determined by λ_j^F 's and λ_j^G 's, and the other is not.

Denote by $\lambda_j = h_j/\bar{H}(a_j-)$ the hazard of H at a_j . Then $(\lambda_1^F, \dots, \lambda_k^F, \lambda_1^G, \dots, \lambda_k^G)$ and $(\lambda_1, \dots, \lambda_k, m(a_1), \dots, m(a_k))$ are one-to-one by

$$(2.2) \quad \lambda_j^F = m(a_j)\lambda_j, \quad \lambda_j^G = \frac{1 - m(a_j)}{1 - m(a_j)\lambda_j}\lambda_j, \quad j=1, \dots, k,$$

where we define $0/0=0$, which occurs in λ_k^G if $m(a_k)=1$. The case of $\lambda_k^F=1$ corresponds to $m(a_k)=1$, and the case of $\lambda_k^G=1$ corresponds to $m(a_k)<1$. Therefore, if $(\lambda_j, m(a_j)), j=1, \dots, k$ are given and $m(a_k)=1$, then $\bar{F}(t)$ is determined for all t and $\bar{G}(t)$ is only for $0 \leq t \leq a_k$. On the other hand, if $m(a_k)<1$, $\bar{G}(t)$ is defined for all t and $\bar{F}(t)$ is only for $0 \leq t \leq a_k$. These results relate to the fact that the KM-estimator has total mass less than one when the largest observation is censored. As we shall see in the next section, an estimated value of $m(a_k)$ is less than one when the largest observation is censored.

2.2. Semi-parametric estimator of the survival function

In view of the censoring pattern, let us consider the KM-estimator (1.5) as maximum likelihood estimator. The likelihood based on the censored observations is given by

$$(2.3) \quad \mathcal{L} = \left[\prod_{j=1}^k \{ \bar{F}(z_{(j)}-) - \bar{F}(z_{(j)}) \}^{d_{1j}} \{ \bar{F}(z_{(j)}) \}^{d_{0j}} \right] \\ \times \left[\prod_{j=1}^k \{ \bar{G}(z_{(j)}-) - \bar{G}(z_{(j)}) \}^{d_{0j}} \{ \bar{G}(z_{(j)}) \}^{d_{1j}} \right] = \mathcal{L}_F \times \mathcal{L}_G.$$

Maximizing $\mathcal{L}_F$ with respect to $\bar{F}$, we obtain the KM-estimator $\bar{F}_n$ defined by (1.5). This is the usual derivation of the KM-estimator, as can be seen in Kalbfleisch and Prentice (1980) Sec. 1.3 and Lawless (1982) Sec. 2.3. We derive it from an idea of the censoring pattern function.

The likelihood $\mathcal{L}$ of (2.3) can be also decomposed as

$$(2.4) \quad \mathcal{L} = \left[\prod_{j=1}^k \{ m(z_{(j)}) \}^{d_{1j}} \{ 1 - m(z_{(j)}) \}^{d_{0j}} \right] \times \left[\prod_{j=1}^k \{ \bar{H}(z_{(j)}-) - \bar{H}(z_{(j)}) \}^{d_j} \right] \\ = \mathcal{L}_m \times \mathcal{L}_H.$$

A survival function which maximizes $\mathcal{L}_H$ is given by the empirical survival function $\bar{H}_n(t)$ defined by (1.3), and which is a discrete survival function with hazard $\hat{\lambda}_j = d_j/n_j$ at $z_{(j)}$, $j=1, 2, \dots, k$. Therefore, a function $\hat{m}(t)$ which

maximizes $\mathcal{L}_m$ is defined at $z_{(j)}$, $j=1, 2, \dots, k$, and $\hat{m}(z_{(j)})=d_{1j}/d_j$. From (2.2), the hazard λ_j^F of F at $z_{(j)}$ is estimated by $\hat{\lambda}_j^F=\hat{m}(z_{(j)})\hat{\lambda}_j=d_{1j}/n_j$. Thus, we obtain the KM-estimator (1.5) from (2.1). If $d_{0k}=0$, then $\hat{m}(z_{(k)})=1$ and hence the KM-estimator $\bar{F}_n(t)$ has total mass one. On the other hand, if $d_{0k}>0$, then $\hat{m}(z_{(k)})<1$ and hence $\bar{F}_n(t)$ is not defined for $t>z_{(k)}$.

In the KM-estimator, the censoring pattern function $m(t)$ was estimated by $\hat{m}(z_{(j)})=d_{1j}/d_j$, $j=1, 2, \dots, k$. If we can assume a parametric form $m(t; \theta)$ with an unknown p -dimensional vector of parameters θ , then $m(z_{(j)}; \theta)$ can be estimated by $m(z_{(j)}; \hat{\theta})$, where $\hat{\theta}$ is defined as θ which maximizes

$$(2.5) \quad \mathcal{L}_m(\theta) = \prod_{j=1}^k \{m(z_{(j)}; \theta)\}^{d_{1j}} \{1 - m(z_{(j)}; \theta)\}^{d_{0j}}.$$

Thus we obtain an estimator

$$(2.6) \quad \bar{F}_n^*(t) = 1 - F_n^*(t) = \prod_{j: z_{(j)} \leq t} \{1 - m(z_{(j)}; \hat{\theta}) d_j / n_j\}.$$

A special case of (2.6) is (1.4), in which the constant censoring pattern is assumed. Another special case of (2.6) is the KM-estimator. Indeed, if we define a k -dimensional vector of parameters $\theta=(\theta_1, \dots, \theta_k)'$ by $\theta_j=m(z_{(j)}; \theta)$, $j=1, \dots, k$, then $\hat{\theta}$ is given by $\hat{\theta}=(d_{11}/d_1, \dots, d_{1k}/d_k)'$ and hence $m(z_{(j)}; \hat{\theta})=d_{1j}/d_j$, $j=1, \dots, k$. By substituting them into (2.6), we have the KM-estimator $\bar{F}_n(t)$ defined by (1.5).

If the assumed parametric censoring pattern function $m(t; \theta)$ is strictly less than one for all t and θ , then the estimator $\bar{F}_n^*(t)$ is always undefined for $t>z_{(k)}$. If $d_{0k}>0$, then the KM-estimator $\bar{F}_n(t)$ is also undefined for $t>z_{(k)}$. Efron (1967) suggests estimating $\bar{F}_n(t)$ by 0 beyond $z_{(k)}$, and Gill (1980) suggests estimating $\bar{F}_n(t)$ by $\bar{F}_n(z_{(k)})$ for $t>z_{(k)}$. Although both tail corrections have the same large-sample properties, a study of small-sample properties by Klein (1991) shows Gill's version is preferred. Thus we define $\bar{F}_n^*(t)$ by $\bar{F}_n^*(z_{(k)})$ for $t>z_{(k)}$, similarly to the Gill's tail correction for the KM-estimator.

2.3. Parametric models for the censoring pattern function

The censoring indicator δ can be regarded as a binary response with a covariate $Z=t$. Therefore, as parametric models of the censoring pattern, we can use ordinary regression models for binary responses such as the logistic regression model. The following models are most common.

Logistic model $m(t; \theta) = \exp\{\eta(t; \theta)\} / [1 + \exp\{\eta(t; \theta)\}]$,

Probit model $m(t; \theta) = \Phi\{\eta(t; \theta)\}$,

Complementary log-log model $m(t; \theta) = 1 - \exp[-\exp\{-\eta(t; \theta)\}]$,

where $\eta(t; \theta)$ is a predictor and Φ is the standard normal distribution function. The most fundamental predictor is a linear predictor $\eta(t; \theta) = \theta_1 + \theta_2 t$, $\theta = (\theta_1, \theta_2)'$. For this linear predictor, the above three models are monotonous in t . More general predictor is a polynomial form $\eta(t; \theta) = \theta_1 + \theta_2 t + \dots + \theta_p t^{p-1}$, $\theta = (\theta_1, \dots, \theta_p)'$, and which gives non-monotonous censoring pattern.

As an example, we assume that F and G are the Weibull distributions given by

$$\bar{F}(t) = \exp(-\alpha_f t^{\beta_f}) \quad \text{and} \quad \bar{G}(t) = \exp(-\alpha_g t^{\beta_g}),$$

where $\alpha_f, \beta_f, \alpha_g$ and β_g are positive. Then the censoring pattern function is given by the logistic model with a predictor $\eta(t; \theta) = \theta_1 + \theta_2 \log t$, where $\theta = (\theta_1, \theta_2)$, $\theta_1 = \log \alpha_f + \log \beta_f - \log \alpha_g - \log \beta_g$ and $\theta_2 = \beta_f - \beta_g$.

2.4. Variance estimation

When the observed times $z_{(1)} < z_{(2)} < \cdots < z_{(k)}$ are fixed and $d_j, j=1, 2, \dots, k$ increase at the same rate as the sample size n , the standard large sample theory for maximum likelihood estimators can be applied.

The likelihood $\mathcal{L}$ of (2.4) is

$$\mathcal{L} = \mathcal{L}_m(\theta) \times \mathcal{L}_H(\lambda_1, \lambda_2, \dots, \lambda_k),$$

where $\mathcal{L}_m(\theta)$ is defined by (2.5) and $\mathcal{L}_H(\lambda_1, \lambda_2, \dots, \lambda_k) = \prod_{j=1}^k \lambda_j^{d_j} (1 - \lambda_j)^{n_j - d_j}$. Thus a covariance matrix of a $(p+k-1)$ -dimensional random vector $(\hat{\theta}', \hat{\lambda}_1, \dots, \hat{\lambda}_{k-1})$ can be estimated by

$$\begin{bmatrix} \hat{I}^{-1}(\hat{\theta}) & 0 \\ 0 & \text{diag}\left(\frac{\hat{\lambda}_1(1-\hat{\lambda}_1)}{n_1}, \dots, \frac{\hat{\lambda}_{k-1}(1-\hat{\lambda}_{k-1})}{n_{k-1}}\right) \end{bmatrix},$$

where

$$\hat{I}(\theta) = -\frac{\partial^2}{\partial \theta \partial \theta'} \log \mathcal{L}_m(\theta).$$

Noting $\hat{\lambda}_k = 1$, from (2.6), we have an expression

$$\log \bar{F}_n^*(t) = \sum_{j=1}^{k-1} I(z_{(j)} \leq t) \log\{1 - m(z_{(j)}; \hat{\theta}) \hat{\lambda}_j\} + I(z_{(k)} \leq t) \log\{1 - m(z_{(k)}; \hat{\theta})\},$$

and which is a function of the $(p+k-1)$ -dimensional vector $(\hat{\theta}', \hat{\lambda}_1, \dots, \hat{\lambda}_{k-1})$. Put $a_{ij}(\theta) = \partial m(z_{(j)}; \theta) / \partial \theta_i$, for $i=1, 2, \dots, p, j=1, 2, \dots, k$, then a covariance matrix of a k -dimensional random vector

$$[\log\{1 - m(z_{(1)}; \hat{\theta}) \hat{\lambda}_1\}, \dots, \log\{1 - m(z_{(k-1)}; \hat{\theta}) \hat{\lambda}_{k-1}\}, \log\{1 - m(z_{(k)}; \hat{\theta})\}]$$

can be estimated by

$$\hat{U}' \hat{I}^{-1}(\hat{\theta}) \hat{U} + \text{diag}\left(\frac{\hat{\lambda}_1(1-\hat{\lambda}_1)\{m(z_{(1)}; \hat{\theta})\}^2}{n_1\{1 - m(z_{(1)}; \hat{\theta}) \hat{\lambda}_1\}^2}, \dots, \frac{\hat{\lambda}_{k-1}(1-\hat{\lambda}_{k-1})\{m(z_{(k-1)}; \hat{\theta})\}^2}{n_{k-1}\{1 - m(z_{(k-1)}; \hat{\theta}) \hat{\lambda}_{k-1}\}^2}, 0\right),$$

where $\hat{U}$ is a $p \times k$ matrix with (i, j) -element $\hat{u}_{ij} = \hat{\lambda}_j a_{ij}(\hat{\theta}) / \{1 - m(z_{(j)}; \hat{\theta}) \hat{\lambda}_j\}$. Hence an asymptotic variance of $\log \bar{F}_n^*(t)$ can be estimated by

$$(2.7) \quad \hat{V}(t) = 1'_k(t) \hat{U}' \hat{I}^{-1}(\hat{\theta}) \hat{U} 1_k(t) + \sum_{j=1}^k I(z_{(j)} \leq t) \frac{\hat{\lambda}_j(1-\hat{\lambda}_j)\{m(z_{(j)}; \hat{\theta})\}^2}{n_j\{1 - m(z_{(j)}; \hat{\theta}) \hat{\lambda}_j\}^2},$$

where $1_k(t)$ is a k -dimensional vector defined by $1_k(t) = [I(z_{(1)} \leq t), \dots, I(z_{(k)} \leq t)]$.

$t)]'$. Thus we obtain an estimated variance

$$(2.8) \quad \text{var}\{\bar{F}_n^*(t)\} = \{\bar{F}_n^*(t)\}^2 \hat{V}(t).$$

In §2.2, we have seen that $\bar{F}_n^*(t)$ coincides with the KM-estimator $\bar{F}_n(t)$ in the non-parametric case in which the parameter vector θ is k -dimensional and $\theta_j = m(z_{(j)}; \theta)$, $j=1, \dots, k$. In this case, $\hat{\theta} = [\hat{\theta}_1, \dots, \hat{\theta}_k]' = [d_{11}/d_1, \dots, d_{1k}/d_k]'$,

$$(2.9) \quad \begin{aligned} \hat{U} &= \text{diag}\left(\frac{\hat{\lambda}_1}{1 - \hat{\theta}_1 \hat{\lambda}_1}, \dots, \frac{\hat{\lambda}_k}{1 - \hat{\theta}_k \hat{\lambda}_k}\right) \quad \text{and} \\ \hat{I}^{-1}(\hat{\theta}) &= \text{diag}\left(\frac{\hat{\theta}_1(1 - \hat{\theta}_1)}{n_1 \hat{\lambda}_1}, \dots, \frac{\hat{\theta}_k(1 - \hat{\theta}_k)}{n_k \hat{\lambda}_k}\right). \end{aligned}$$

From (2.7), (2.8) and (2.9), we have

$$\begin{aligned} \hat{V}(t) &= \sum_{j=1}^k I(z_{(j)} \leq t) \frac{\hat{\theta}_j(1 - \hat{\theta}_j) \hat{\lambda}_j}{n_j(1 - \hat{\theta}_j \hat{\lambda}_j)^2} + \sum_{j=1}^k I(z_{(j)} \leq t) \frac{\hat{\lambda}_j(1 - \hat{\lambda}_j) \hat{\theta}_j^2}{n_j(1 - \hat{\theta}_j \hat{\lambda}_j)^2} \\ &= \sum_{j=1}^k I(z_{(j)} \leq t) \frac{\hat{\theta}_j \hat{\lambda}_j}{n_j(1 - \hat{\theta}_j \hat{\lambda}_j)} = \sum_{j=1}^k I(z_{(j)} \leq t) \frac{d_{1j}}{n_j(n_j - d_{1j})} \end{aligned}$$

and

$$\text{var}\{\bar{F}_n^*(t)\} = \{\bar{F}_n^*(t)\}^2 \sum_{j=1}^k I(z_{(j)} \leq t) \frac{d_{1j}}{n_j(n_j - d_{1j})}.$$

This is the well-known Greenwood's formula. Thus the variance estimator (2.8) can be regarded as a generalization of the Greenwood's formula.

2.5. Estimation of cumulative hazard function

The cumulative hazard function Λ^F corresponding to the distribution F is defined by

$$(2.10) \quad \Lambda^F(t) = \int_0^t \frac{dF(s)}{\bar{F}(s-)}.$$

Throughout this paper, $\int_0^t$ means integration on a closed interval $[0, t]$. Estimating Λ^F is important in selecting between parametric models for F and providing crude estimate of the hazard rate of F . By substituting the KM-estimator $\bar{F}_n$ into (2.10), we have

$$\Lambda_n^F(t) = \int_0^t \frac{dF_n(s)}{\bar{F}_n(s-)} = \sum_{j=1}^k I(z_{(j)} \leq t) \frac{d_{1j}}{n_j},$$

and which was first suggested by Nelson (1972) and rediscovered by Aalen (1978) using counting process techniques. This estimator is referred to as the Nelson-Aalen estimator (NA-estimator). The estimated variances of the NA-estimator are due to Aalen (1978) and Klein (1991), and which are given by

$$(2.11) \quad \text{var}\{\Lambda_n^F(t)\} = \begin{cases} \sum_{j=1}^k I(z_{(j)} \leq t) d_{1j} / n_j^2 & \text{by Aalen,} \\ \sum_{j=1}^k I(z_{(j)} \leq t) d_{1j} (n_j - d_{1j}) / n_j^3 & \text{by Klein.} \end{cases}$$

By substituting the estimator $\bar{F}_n^*$ into (2.10), we obtain an estimator

$$\Lambda_n^{F*}(t) = \sum_{j=1}^k I(z_{(j)} \leq t) \frac{m(z_{(j)}; \hat{\theta}) d_j}{n_j}.$$

Similarly to the variance of the estimator $\bar{F}_n^*$, the estimated variance of $\Lambda_n^{F*}(t)$ can be derived as

$$(2.12) \quad \text{var}\{\Lambda_n^{F*}(t)\} = 1'_k(t) \hat{W}' \hat{I}^{-1}(\hat{\theta}) \hat{W} 1_k(t) + \sum_{j=1}^k I(z_{(j)} \leq t) \frac{\hat{\lambda}_j(1 - \hat{\lambda}_j) \{m(z_{(j)}; \hat{\theta})\}^2}{n_j},$$

where $\hat{W}$ is a $p \times k$ matrix with (i, j) -element $\hat{w}_{ij} = \hat{\lambda}_j a_{ij}(\hat{\theta})$.

In the nonparametric case in which the parameter vector θ is k -dimensional and $\theta_j = m(z_{(j)}; \theta)$, $j=1, \dots, k$, the estimator Λ_n^{F*} reduces to the NA-estimator. In this case $\hat{I}^{-1}(\hat{\theta})$ is given by (2.9) and $\hat{W} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_k)$, and hence we can see that the formula (2.12) reduces to the estimated variance (2.11) by Klein (1991).

The estimated variances (2.8) and (2.12) will be justified in the next section.

3. Asymptotic properties

In this section we investigate large sample properties of the estimators F_n^* and Λ_n^{F*} , and discuss comparison of them with the KM-estimator and NA-estimator. Apart from technical regularity conditions, the maximum likelihood estimator $\hat{\theta}$ converges to θ_0 in probability as $n \rightarrow \infty$ and $n^{1/2}(\hat{\theta} - \theta_0)$ is asymptotically normal with mean zero and covariance matrix $I^{-1}(\theta_0)$, where θ_0 is the true vector of parameters, $I(\theta) = E\{B(Z; \theta)B'(Z; \theta)\}$ and $B(Z; \theta)$ is a p -dimensional random vector defined by

$$B(Z; \theta) = \{m(Z; \theta)\}^{-\frac{1}{2}} \{1 - m(Z; \theta)\}^{-\frac{1}{2}} \{\partial m(Z; \theta) / \partial \theta\}.$$

The following theorem, proved in the Appendix, states that the estimators F_n^* and Λ_n^{F*} are uniformly consistent estimators of F and Λ^F on interval $[0, t]$ for which $F(t) < 1$ and $0 \leq t < \tau$, respectively, where $\tau = \sup\{z: \bar{H}(z) > 0\}$.

THEOREM 1. *If $t \in [0, \tau)$ is such that $F(t) < 1$, then*

$$(3.1) \quad \sup_{0 \leq s \leq t} |\Lambda_n^{F*}(s) - \Lambda^F(s)| \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty$$

and

$$(3.2) \quad \sup_{0 \leq s \leq t} |\bar{F}_n^*(s) - \bar{F}(s)| \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty,$$

where $\xrightarrow{P}$ means convergence in probability.

It is well-known that the NA-estimator Λ_n^F and the KM-estimator $\bar{F}_n$ are also uniformly consistent (Gill, 1983; Wang, 1987). So it is important to investigate the asymptotic relative efficiency of Λ_n^{F*} respect to Λ_n^F and that of $\bar{F}_n^*$ respect to $\bar{F}_n$. It is known that $n^{1/2}\{\Lambda_n^F(t) - \Lambda^F(t)\}$ is asymptotically normal with mean zero and variance

$$(3.3) \quad \sigma^2(t) = \int_0^t \frac{1 - \Delta\Lambda^F(u)}{\bar{H}(u-)} d\Lambda^F(u) = \int_0^t \frac{m(u; \theta_0)\{1 - m(u; \theta_0)\Delta\Lambda(u)\}}{\bar{H}(u-)} d\Lambda(u)$$

and $n^{1/2}\{\bar{F}_n(t) - \bar{F}(t)\}$ is asymptotically normal with mean zero and variance $v^2(t)$, where

$$(3.4) \quad \begin{aligned} \frac{v^2(t)}{\{\bar{F}(t)\}^2} &= \int_0^t \frac{\{\bar{F}(u-)\}^2\{1 - \Delta\Lambda^F(u)\}}{\{\bar{F}(u)\}^2\bar{H}(u-)} d\Lambda^F(u) \\ &= \int_0^t \frac{m(u; \theta_0)}{\bar{H}(u-)\{1 - m(u; \theta_0)\Delta\Lambda(u)\}} d\Lambda(u), \end{aligned}$$

Λ is the cumulative hazard function corresponding to the distribution H , $\Delta\Lambda^F(t) = \Lambda^F(t) - \Lambda^F(t-)$ and $\Delta\Lambda(t) = \Lambda(t) - \Lambda(t-)$. On the other hand, asymptotic variances of the estimators $\Lambda_n^{F*}(t)$ and $\bar{F}_n^*(t)$ are given by the next theorem proved in the Appendix.

THEOREM 2. Assume that F and G do not have common jumps. Then, for $t \in [0, \tau)$ such that $F(t) < 1$, the asymptotic distribution of $n^{1/2}\{\Lambda_n^{F*}(t) - \Lambda^F(t)\}$ is normal with mean zero and variance $\sigma^{*2}(t)$, and the asymptotic distribution of $n^{1/2}\{\bar{F}_n^*(t) - \bar{F}(t)\}$ is normal with mean zero and variance $v^{*2}(t)$, where

$$(3.5) \quad \sigma^{*2}(t) = W'(t; \theta_0)I^{-1}(\theta_0)W(t; \theta_0) + \int_0^t \frac{\{m(u; \theta_0)\}^2\{1 - \Delta\Lambda(u)\}}{\bar{H}(u-)} d\Lambda(u),$$

$$(3.6) \quad \begin{aligned} \frac{v^{*2}(t)}{\{\bar{F}(t)\}^2} &= U'(t; \theta_0)I^{-1}(\theta_0)U(t; \theta_0) \\ &\quad + \int_0^t \frac{\{m(u; \theta_0)\}^2\{1 - \Delta\Lambda(u)\}}{\bar{H}(u-)\{1 - m(u; \theta_0)\Delta\Lambda(u)\}^2} d\Lambda(u), \end{aligned}$$

$$W(t; \theta) = \int_0^t \frac{\partial m(u; \theta)}{\partial \theta} d\Lambda(u) \text{ and}$$

$$U(t; \theta) = \int_0^t \frac{\partial m(u; \theta)}{\partial \theta} \{1 - m(u; \theta)\Delta\Lambda(u)\}^{-1} d\Lambda(u).$$

Define

$$\bar{W}_n(t; \theta) = \int_0^t \frac{\partial m(u; \theta)}{\partial \theta} d\Lambda_n(u)$$

and

$$\bar{U}_n(t; \theta) = \int_0^t \frac{\partial m(u; \theta)}{\partial \theta} \{1 - m(u; \theta)\Delta\Lambda_n(u)\}^{-1} d\Lambda_n(u),$$

where $\Lambda_n(t) = \int_0^t dH_n(u)/\bar{H}_n(u-)$ and $\Delta\Lambda_n(t) = \Lambda_n(t) - \Lambda_n(t-)$. Then we can

express estimated variances (2.8) and (2.12) as

$$\frac{n \times (2.8)}{\{\bar{F}_n^*(t)\}^2} = \hat{U}_n'(t; \hat{\theta}) \{\hat{I}(\hat{\theta})/n\}^{-1} \hat{U}_n(t; \hat{\theta}) + \int_0^t \frac{\{m(u; \hat{\theta})\}^2 \{1 - \Delta \Lambda_n(u)\}}{\bar{H}_n(u-)\{1 - m(u; \hat{\theta}) \Delta \Lambda_n(u)\}^2} d\Lambda_n(u)$$

and

$$n \times (2.12) = \hat{W}_n'(t; \hat{\theta}) \{\hat{I}(\hat{\theta})/n\}^{-1} \hat{W}_n(t; \hat{\theta}) + \int_0^t \frac{\{m(u; \hat{\theta})\}^2 \{1 - \Delta \Lambda_n(u)\}}{\bar{H}_n(u-)} d\Lambda_n(u).$$

Thus, it can be verified that the estimated variances $n \times (2.8)$ and $n \times (2.12)$ are consistent estimators of $v^{*2}(t)$ and $\sigma^{*2}(t)$, respectively.

Denote the asymptotic relative efficiency of the NA-estimator $\Lambda_n^F(t)$ respect to the estimator $\Lambda_n^{F*}(t)$ by $\text{ARE}\{\Lambda_n^F(t), \Lambda_n^{F*}(t)\}$ and that of the KM-estimator $\bar{F}_n(t)$ respect to the estimator $\bar{F}_n^*(t)$ by $\text{ARE}\{\bar{F}_n(t), \bar{F}_n^*(t)\}$. Then, from (3.3), (3.4), (3.5) and (3.6), we have expressions

$$(3.7) \quad \sigma^2(t) [\text{ARE}\{\Lambda_n^F(t), \Lambda_n^{F*}(t)\} - 1] \\ = E\{\gamma(Z; t, \theta_0) B'(Z; \theta_0)\} I^{-1}(\theta_0) E\{\gamma(Z; t, \theta_0) B(Z; \theta_0)\} \\ - E[\{\gamma(Z; t, \theta_0)\}^2]$$

and

$$(3.8) \quad v^2(t) \{\bar{F}(t)\}^{-2} [\text{ARE}\{\bar{F}_n(t), \bar{F}_n^*(t)\} - 1] \\ = E\{\xi(Z; t, \theta_0) B'(Z; \theta_0)\} I^{-1}(\theta_0) E\{\xi(Z; t, \theta_0) B(Z; \theta_0)\} \\ - E[\{\xi(Z; t, \theta_0)\}^2],$$

where

$$\gamma(Z; t, \theta_0) = \{m(Z; \theta_0)\}^{\frac{1}{2}} \{1 - m(Z; \theta_0)\}^{\frac{1}{2}} I(Z \leq t) \{\bar{H}(Z-)\}^{-1}$$

and

$$\xi(Z; t, \theta_0) = \{m(Z; \theta_0)\}^{\frac{1}{2}} \{1 - m(Z; \theta_0)\}^{\frac{1}{2}} I(Z \leq t) \{\bar{F}(Z-)\} \{\bar{F}(Z)\}^{-1} \{\bar{H}(Z-)\}^{-1}.$$

The right-hand sides of (3.7) and (3.8) are not positive since

$$\frac{E\{\gamma(Z; t, \theta_0) B'(Z; \theta_0)\} I^{-1}(\theta_0) E\{\gamma(Z; t, \theta_0) B(Z; \theta_0)\}}{E[\{\gamma(Z; t, \theta_0)\}^2]}$$

and

$$\frac{E\{\xi(Z; t, \theta_0) B'(Z; \theta_0)\} I^{-1}(\theta_0) E\{\xi(Z; t, \theta_0) B(Z; \theta_0)\}}{E[\{\xi(Z; t, \theta_0)\}^2]}$$

are multiple correlation coefficients between $\gamma(Z; t, \theta_0)$ and $B(Z; \theta_0)$ and between $\xi(Z; t, \theta_0)$ and $B(Z; \theta_0)$, respectively. Thus we can see that the estimator $\bar{F}_n^*(t)$ ($\Lambda_n^{F*}(t)$) is asymptotically more efficient than the KM-estimator $\bar{F}_n(t)$ (NA-estimator $\Lambda_n^F(t)$).

As an example, we assume that $\bar{F}$ and $\bar{G}$ are exponential distributions given by $\bar{F}(x) = e^{-x}$ and $\bar{G}(y) = e^{-\lambda c y}$. When the constant model of censoring pattern $m(t; \theta) = \theta$ is used, from (3.7) and (3.8), we have

$$(3.9) \quad \text{ARE}\{\Lambda_n^E(t), \Lambda_n^{F*}(t)\} = \text{ARE}\{\bar{F}_n(t), \bar{F}_n^*(t)\} \\ = \frac{1}{1+\lambda^G} + \frac{\lambda^G(1+\lambda^G)t^2}{\exp\{(1+\lambda^G)t\}-1}.$$

Denote (3.9) by $\text{ARE}(t)$, then $\text{ARE}(t) < 1$ for any t . The function $\text{ARE}(t)$ has a maximum $(1+a\lambda^G)/(1+\lambda^G)$ at $t=b/(1+\lambda^G)$, where $a=b^2(e^b-1)^{-1} \approx 0.601$ and $b \approx 2.218$ is a solution of an equation $(b-2)e^b=2$. For $\lambda^G=0.5, 1.0$ and 2.0 , the maximum values are $(2+a)/3 \approx 0.867$, $(1+a)/2 \approx 0.800$ and $(1+2a)/3 \approx 0.734$, respectively. From these results, we can see that the estimator $\bar{F}_n^*$ is more efficient than the KM-estimator for heavy censored data when the sample size is large. In §5, by a simulation study, small sample comparison is discussed in the same situation as this example.

4. An example

As a numerical example, we consider the leukemia data of Table 1. Partition a time interval $[5, 40)$ as $[5, 10)$, $[10, 15)$, $\dots$, $[35, 40)$. Proportions of uncensored observations in the intervals are $4/6, 2/4, 1/3, 2/3, 0/1, 0/3$ and $0/1$. As a representative of each interval, we choose the midpoint, and seven points

Figure 1. Comparison of the fitted censoring pattern functions based on the model (4.2) for the leukemia data of Table 1. The plotted points are points of (4.1). The line - - - is the constant fitting, the curve — is the linear logistic fitting, the curve --- is the quadratic logistic fitting, and the curve ---- is the cubic logistic fitting.

$$(4.1) \quad (7.5, 4/6), (12.5, 2/4), (17.5, 1/3), (22.5, 2/3), (27.5, 0/1), (32.5, 0/3), (37.5, 0/1)$$

are plotted in Figure 1. From this plot, we can perceive an outline of the censoring pattern of this data.

For the censoring pattern, we assume a logistic model with a predictor of a cubic polynomial form:

$$(4.2) \quad m(t; \theta) = \frac{\exp\{\eta(t; \theta)\}}{1 + \exp\{\eta(t; \theta)\}}, \quad \eta(t; \theta) = \sum_{i=1}^4 \theta_i (t - \bar{t})^{i-1}, \quad \theta = (\theta_1, \dots, \theta_4)',$$

where $\bar{t} = 17.095$ is the sample mean of the observed times. When $\theta_2 = \theta_3 = \theta_4 = 0$, this is the model of the constant censoring pattern. If $\theta_3 = \theta_4 = 0$, this is the linear logstic model. If $\theta_4 = 0$, this is the quadratic logistic model. For each model, estimated values of θ_i 's are in Table 2, and estimated censoring pattern functions are shown in Figure 1.

To compare these models, we calculated values of the information criteria $AIC = -2 \log \mathcal{L}_m(\hat{\theta}) + 2p$ (Akaike, 1973), $AICC = -2 \log \mathcal{L}_m(\hat{\theta}) + 2pn/(n-p-1)$

Table 2. The estimated parameters and the model selection criteria in the logistic censoring pattern model (4.2) for the leukemia data of Table 1

Model	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	AIC	AICC	BIC	LR-test (p-value)
Constant	-0.288	0	0	0	30.68	30.89	31.73	6.791 (0.079)
Linear	-0.410	-0.112	0	0	28.13	28.80	30.22	2.240 (0.326)
Quadratic	-0.107	-0.109	-0.004	0	29.80	31.22	32.94	1.913 (0.167)
Cubic	-0.188	0.108	-0.004	-0.002	29.89	32.39	34.07	—

Table 3. Construction of the KM-estimator $\bar{F}_n$ and the proposed semi-parametric estimator $\bar{F}_n^*$ based on the linear logistic model of the censoring pattern for the leukemia data of Table 1

$z_{(j)}$	d_j	d_{1j}	n_j	$1 - \frac{d_{1j}}{n_j}$	$1 - \frac{d_j m(z_{(j)}; \hat{\theta})}{n_j}$	$\bar{F}_n(z_{(j)})$	$\bar{F}_n^*(z_{(j)})$
6	4	3	21	0.8571	0.8671	0.8571	0.8671
7	1	1	17	0.9412	0.9604	0.8067	0.8328
9	1	0	16	1.0000	0.9611	0.8067	0.8004
10	2	1	15	0.9333	0.9206	0.7529	0.7369
11	1	0	13	1.0000	0.9563	0.7529	0.7046
13	1	1	12	0.9167	0.9573	0.6902	0.6746
16	1	1	11	0.9091	0.9610	0.6275	0.6483
17	1	0	10	1.0000	0.9599	0.6275	0.6222
19	1	0	9	1.0000	0.9612	0.6275	0.5981
20	1	0	8	1.0000	0.9595	0.6275	0.5739
22	1	1	7	0.8571	0.9605	0.5378	0.5512
23	1	1	6	0.8333	0.9575	0.4482	0.5278
25	1	0	5	1.0000	0.9571	0.4482	0.5052
32	2	0	4	1.0000	0.9446	0.4482	0.4772
34	1	0	2	1.0000	0.9548	0.4482	0.4556
35	1	0	1	1.0000	0.9184	0.4482	0.4184

Figure 2. Comparison of survival curves for the leukemia data of Table 1. The dotted step function is the KM-estimator, and the solid step function is the proposed semi-parametric estimator based on the linear logistic model of the censoring pattern.

Table 4. Estimated standard errors of the KM-estimator $\bar{F}_n$ and the proposed estimator $\bar{F}_n^*$ based on the linear logistic censoring pattern for the leukemia data of Table 1

Time on study	Survival Function		Standard Error (SE)	
(t)	$\bar{F}_n(t)$	$\bar{F}_n^*(t)$	$SE\{\bar{F}_n(t)\}$	$SE\{\bar{F}_n^*(t)\}$
$0 \leq t < 6$	1.0000	1.0000	0.000	0.000
$6 \leq t < 7$	0.8571	0.8671	0.076	0.066
$7 \leq t < 9$	0.8067	0.8328	0.087	0.075
$9 \leq t < 10$	0.8067	0.8004	0.087	0.082
$10 \leq t < 11$	0.7529	0.7369	0.096	0.094
$11 \leq t < 13$	0.7529	0.7046	0.096	0.099
$13 \leq t < 16$	0.6902	0.6746	0.107	0.103
$16 \leq t < 17$	0.6275	0.6483	0.114	0.105
$17 \leq t < 19$	0.6275	0.6222	0.114	0.108
$19 \leq t < 20$	0.6275	0.5981	0.114	0.109
$20 \leq t < 22$	0.6275	0.5739	0.114	0.111
$22 \leq t < 23$	0.5378	0.5512	0.128	0.113
$23 \leq t < 25$	0.4482	0.5278	0.135	0.116
$25 \leq t < 32$	0.4482	0.5052	0.135	0.120
$32 \leq t < 34$	0.4482	0.4772	0.135	0.128
$34 \leq t < 35$	0.4482	0.4556	0.135	0.138

(Hurvich and Tsai, 1989) and $BIC = -2 \log \mathcal{L}_m(\hat{\theta}) + p \log n$ (Schwarz, 1978), where $\mathcal{L}_m(\theta)$ is defined in (2.5) and p is the number of unknown parameters; $p=1$ (constant), $p=2$ (linear), $p=3$ (quadratic) and $p=4$ (cubic). We also performed likelihood ratio test of a null hypothesis that the censoring pattern is constant (linear logistic, quadratic logistic) against a hypothesis that it is cubic logistic. The results are shown in Table 2. From these results, we can see that the linear logistic model is best among these models.

The calculations needed to construct the KM-estimator $\bar{F}_n$ and the estimator $\bar{F}_n^*$ based on the linear logistic model of the censoring pattern are in Table 3. Figure 2 shows plots of these estimated survival functions. From Figure 2, we can see that the estimator $\bar{F}_n^*$ provide more detailed information about survival probabilities than the KM-estimator in time intervals in which no remission time is observed, for example, time intervals (16, 22) and (23, 35).

The estimated variances of these estimators are in Table 4. The standard errors of the estimators $\bar{F}_n^*$ are almost smaller than that of the KM-estimator.

5. Simulation results

In this section we report on a simulation study which is likely to demonstrate the impact of the survival function estimator based on modeling of the censoring pattern.

We assume that the lifetime X and the censoring time Y are exponentially distributed: $\bar{F}(x) = e^{-x}$ and $\bar{G}(y) = e^{-\lambda^G y}$, with varying λ^G 's. An assumed parametric model of the censoring pattern is the constant model: $m(t) = \text{constant}$. This is the case in which the censoring pattern is correctly specified. We compare the estimator $\bar{F}_n^*$ with the KM-estimator under correct specification of the censoring pattern.

Sample size are $n=10, 30$ and 60 . Three values of λ^G were chosen so that probabilities of censoring are $1/3, 1/2$ and $2/3$. For each n and every λ^G , 300000 samples were drawn, and mean absolute errors (MAE)

$$MAE^{KM} = E\{|\bar{F}_n(t_a) - \bar{F}(t_a)|\} \quad \text{and} \quad MAE^* = E\{|\bar{F}_n^*(t_a) - \bar{F}(t_a)|\}$$

were computed, where t_a is defined by $\bar{F}(t_a) = \exp(-t_a) = \alpha$ for $\alpha = 0.90, 0.75, 0.50, 0.25$ and 0.10 . The results are shown in Table 5.

Table 5. Simulation results for mean absolute errors (MAE) of the KM-estimator $\bar{F}_n$ and the proposed estimator $\bar{F}_n^*$ under constant censoring pattern; MAE^{KM} is MAE of the KM-estimator, and MAE^* is MAE of $\bar{F}_n^*$. The true survival probability is α

		$n=10$			$n=30$			$n=60$		
α	λ^G	0.50	1.00	2.00	0.50	1.00	2.00	0.50	1.00	2.00
0.90	MAE^*/MAE^{KM}	0.912	0.817	0.742	0.859	0.778	0.708	0.848	0.770	0.715
0.75	MAE^*/MAE^{KM}	0.869	0.838	0.820	0.881	0.840	0.830	0.882	0.842	0.831
0.50	MAE^*/MAE^{KM}	0.916	0.879	0.804	0.924	0.896	0.838	0.927	0.904	0.851
0.25	MAE^*/MAE^{KM}	0.898	0.774	0.715	0.922	0.827	0.613	0.930	0.852	0.622
0.10	MAE^*/MAE^{KM}	0.720	0.744	0.915	0.827	0.586	0.770	0.864	0.622	0.661

It follows from Table 5 that the estimator $\bar{F}_n^*$ always has smaller MAE than the KM-estimator. The reduction of MAE is substantial for $\lambda^G \geq 1$ and $\alpha \geq 0.25$. However when $n=10$, $\lambda^G \geq 2$ and $\alpha=0.10$, the superiority of $\bar{F}_n^*$ is slight. From these results, we can say that, the estimator $\bar{F}_n^*$ is generally preferable than the KM-estimator, particularly for heavy censored and large sample data, however estimation of the right-tail probability is not so improved in the case of heavy censoring and small sample size.

We should notice that the simulation results reported in this section and the theoretical results obtained in §3 are based on correct specification of the censoring pattern function. However it may be difficult to know exactly the censoring pattern function in most cases. Thus it is important to consider the sensitivity or robustness of the semi-parametric estimator with respect to the model of censoring pattern. There is room for further investigation on this point.

Appendix

Define $\Lambda(t) = \int_0^t dH(u)/\bar{H}(u-)$, $\Lambda_n(t) = \int_0^t dH_n(u)/\bar{H}_n(u-)$, $R(t) = \sum_{i=1}^n I(Z_i \geq t)$, $N(t) = \sum_{i=1}^n I(Z_i \leq t)$, $M(t) = N(t) - \int_0^t R(u) d\Lambda(u)$ and $T_n = \inf\{t: R(t) = 0\}$, and denote by $\bar{F}_{n0}^* = 1 - F_{n0}^*$ the survival function obtained by replacing $\hat{\theta}$ in $\bar{F}_n^*$ by the true vector of parameters θ_0 .

LEMMA 1. For any $t \in [0, \infty)$ such that $\bar{F}(t) > 0$, it holds that

$$\begin{aligned} \bar{F}_n^*(t) - \bar{F}(t) = & -\bar{F}_{n0}^*(t) \int_0^t \{\bar{F}_{n0}^*(u)\}^{-1} \bar{F}_n^*(u-) \{m(u; \hat{\theta}) - m(u; \theta_0)\} d\Lambda_n(u) \\ & - \bar{F}(t) \int_0^t \{\bar{F}(u)\}^{-1} \bar{F}_{n0}^*(u-) m(u; \theta_0) \{R(u)\}^{-1} I(R(u) > 0) dM(u) \\ & + I(T_n < t) \bar{F}_{n0}^*(T_n) \{\bar{F}(T_n)\}^{-1} \{\bar{F}(T_n) - \bar{F}(t)\}. \end{aligned}$$

PROOF of LEMMA 1. We decompose as

$$\bar{F}_n^*(t) - \bar{F}(t) = \{\bar{F}_n^*(t) - \bar{F}_{n0}^*(t)\} + \{\bar{F}_{n0}^*(t) - \bar{F}(t)\}.$$

Let $\Lambda_{n0}^*(t) = \int_0^t dF_{n0}^*(u)/\bar{F}_{n0}^*(u-)$. By Theorem A.1.2 in Appendix of Fleming & Harrington (1991), we have

$$\begin{aligned} \frac{\bar{F}_n^*(t) - \bar{F}_{n0}^*(t)}{\bar{F}_{n0}^*(t)} = & - \int_0^t \{\bar{F}_{n0}^*(u)\}^{-1} \bar{F}_n^*(u-) \{d\Lambda_{n0}^*(u) - d\Lambda_n^*(u)\} \\ = & - \int_0^t \{\bar{F}_{n0}^*(u)\}^{-1} \bar{F}_n^*(u-) \{m(u; \hat{\theta}) - m(u; \theta_0)\} d\Lambda_n(u) \end{aligned}$$

and

$$\begin{aligned} \frac{\bar{F}_{n0}^*(t) - \bar{F}(t)}{\bar{F}(t)} = & - \int_0^t \{\bar{F}(u)\}^{-1} \bar{F}_{n0}^*(u-) \{d\Lambda_{n0}^*(u) - d\Lambda^F(u)\} \\ = & - \int_0^t \{\bar{F}(u)\}^{-1} \bar{F}_{n0}^*(u-) m(u; \theta_0) \{d\Lambda_n(u) - d\Lambda(u)\} \end{aligned}$$

$$= - \int_0^t \{\bar{F}(u)\}^{-1} \bar{F}_{n0}^*(u-) m(u; \theta_0) \{R(u)\}^{-1} I(R(u) > 0) dM(u) \\ + I(T_n < t) \bar{F}_{n0}^*(T_n) \{\bar{F}(T_n)\}^{-1} \{\bar{F}(t)\}^{-1} \{\bar{F}(T_n) - \bar{F}(t)\}.$$

Substituting them into the above decomposition, we have the result. $\square$

PROOF of THEOREM 1. For $s \in [0, t]$, it holds that

$$|\Lambda_n^{F*}(s) - \Lambda^F(s)| \\ \leq \left| \int_0^s \frac{m(u; \theta_0) I(R(u) > 0)}{R(u)} dM(u) \right| + \int_0^s \frac{I(R(u) > 0) |m(u; \hat{\theta}) - m(u; \theta_0)|}{R(u)} dN(u) \\ + I(R(t) = 0) \Lambda(t).$$

From this, we have

$$(A.1) \quad \sup_{0 \leq s \leq t} |\Lambda_n^{F*}(s) - \Lambda^F(s)| \leq \sup_{0 \leq s \leq t} \left| \int_0^s \frac{m(u; \theta_0) I(R(u) > 0)}{R(u)} dM(u) \right| \\ + I(R(t) = 0) \Lambda(t) \\ + \int_0^t \frac{I(R(u) > 0) |m(u; \hat{\theta}) - m(u; \theta_0)|}{R(u)} dN(u).$$

The third term on the right-hand side of (A.1) converges to zero in probability as $n \rightarrow \infty$ by consistency of $\hat{\theta}$, and the second term converges to zero since $H(t) < 1$. So to show (3.1), it suffices to show

$$(A.2) \quad \sup_{0 \leq s \leq t} \left\{ \int_0^s \frac{m(u; \theta_0) I(R(u) > 0)}{R(u)} dM(u) \right\}^2 \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

By Corollary 3.4.1 of Fleming and Harrington (1991) to Lenglar (1977)'s inequality, for any $\varepsilon, \eta > 0$,

$$\text{pr} \left[\sup_{0 \leq s \leq t} \left\{ \int_0^s \frac{m(u; \theta_0) I(R(u) > 0)}{R(u)} dM(u) \right\}^2 \geq \varepsilon \right] \\ \leq \frac{\eta}{\varepsilon} + \text{pr} \left[\int_0^t \frac{\{m(u; \theta_0)\}^2 I(R(u) > 0) \{1 - \Delta \Lambda(u)\}}{R(u)} d\Lambda(u) \geq \eta \right] \\ \leq \frac{\eta}{\varepsilon} + \text{pr} \left[\int_0^t \frac{I(R(u) > 0) \{1 - \Delta \Lambda(u)\}}{R(u)} d\Lambda(u) \geq \eta \right] \\ \leq \frac{\eta}{\varepsilon} + \text{pr} \left[\frac{I(R(t) > 0)}{R(t)} \int_0^t \{1 - \Delta \Lambda(u)\} d\Lambda(u) \geq \eta \right].$$

The second term on the right-hand side converges to zero as $n \rightarrow \infty$ since $R(t) \rightarrow \infty$ as $n \rightarrow \infty$. Since ε and η are arbitrary, (A.2) is established. We next show that (3.2) holds. From Lemma 1, it follows that

$$(A.3) \quad \sup_{0 \leq s \leq t} |\bar{F}_n^*(s) - \bar{F}(s)| \leq \sup_{0 \leq s \leq t} \left| \int_0^s \frac{\bar{F}_{n0}^*(u-) m(u; \theta_0) I(R(u) > 0)}{\bar{F}(u) R(u)} dM(u) \right| \\ + \int_0^t \{\bar{F}_{n0}^*(u)\}^{-1} \bar{F}_n^*(u-) |m(u; \hat{\theta}) - m(u; \theta_0)| d\Lambda_n(u) \\ + I(T_n < t) \bar{F}_{n0}^*(T_n) \{\bar{F}(T_n)\}^{-1} \{\bar{F}(T_n) - \bar{F}(t)\}.$$

The second term on the right-hand side of (A.3) converge to zero in probability as $n \rightarrow \infty$ by consistency of $\hat{\theta}$, and the third term converges to zero as $n \rightarrow \infty$ since $I(T_n < t) \xrightarrow{P} 0$. It also follows, from the Lengart's inequality, the first term converges to zero in probability as $n \rightarrow \infty$. $\square$

PROOF of THEOREM 2. Decompose $n^{1/2}\{\Lambda_n^*(t) - \Lambda^F(t)\}$ as

$$(A.4) \quad \begin{aligned} n^{\frac{1}{2}}\{\Lambda_n^*(t) - \Lambda^F(t)\} \\ = n^{\frac{1}{2}} \int_0^t m(u; \theta_0) I(R(u) > 0) \{R(u)\}^{-1} dM(u) \\ + W_n'(t; \tilde{\theta}) \{n^{\frac{1}{2}}(\tilde{\theta}_n - \theta_0)\} + n^{\frac{1}{2}} \int_0^t I(R(u) = 0) d\Lambda^F(u), \end{aligned}$$

where $W_n(t; \theta)$ is a p -dimensional vector defined by

$$W_n(t; \theta) = \int_0^t \frac{I(R(u) > 0)}{R(u)} \frac{\partial m(u; \theta)}{\partial \theta} dN(u)$$

and $\tilde{\theta}$ is on the line segment between $\hat{\theta}$ and θ_0 . By martingale method, we can see that the asymptotic distribution of the first term on the right-hand side of (A.4) is normal with mean zero and variance given by the second term on the right-hand side of (3.5). It is also obvious that the third term on the right-hand side of (A.4) converges to zero in probability as $n \rightarrow \infty$. Since $W_n(t; \tilde{\theta})$ converges to $W(t; \theta_0)$, the asymptotic distribution of the second term of (A.4) is normal with mean zero and variance given by the first term of (3.5). Since $\hat{\theta}$ is a function of a score vector $S_n(\theta) = \partial \log \mathcal{L}_n(\theta) / \partial \theta$, it would be sufficient to show that the first term on the right-hand side of (A.4) is independent of the score vector. The score vector $S_n(\theta)$ is the value at $t = \infty$ of the process

$$S_n(\theta, t) = \int_0^t \{m(s; \theta)\}^{-1} \frac{\partial m(s; \theta)}{\partial \theta} dM_1(s; \theta) - \int_0^t \{1 - m(s; \theta)\}^{-1} \frac{\partial m(s; \theta)}{\partial \theta} dM_0(s; \theta),$$

where

$$M_1(s; \theta) = \sum_{i=1}^n I(Z_i \leq s, \delta_i = 1) - \int_0^s R(u) m(u; \theta) d\Lambda(u)$$

and

$$M_0(s; \theta) = \sum_{i=1}^n I(Z_i \leq s, \delta_i = 0) - \int_0^s R(u) \{1 - m(u; \theta)\} d\Lambda(u).$$

Denote the first term on the right-hand side of (A.4) by $A_n(t)$. Then we have

$$\begin{aligned} \langle S_n(\theta_0, \cdot), A_n \rangle(t) &= \int_0^t n^{\frac{1}{2}} \left\{ \frac{\partial m(s; \theta_0)}{\partial \theta} \right\} \frac{I(R(s) > 0)}{R(s)} d \langle M_1(\cdot; \theta_0), M \rangle(s) \\ &\quad - \int_0^t n^{\frac{1}{2}} \left\{ \frac{\partial m(s; \theta_0)}{\partial \theta} \right\} \frac{m(s; \theta_0)}{1 - m(s; \theta_0)} \frac{I(R(s) > 0)}{R(s)} d \langle M_0(\cdot; \theta_0), M \rangle(s) \\ &= 0, \end{aligned}$$

and hence the processes $S_n(\theta_0, t)$ and $A_n(t)$ are orthogonal. We next consider

the asymptotic distribution of $n^{1/2}\{F_n^*(t) - F(t)\}$. From Lemma 1, we have an expression

$$(A.5) \quad n^{1/2}\{\bar{F}_n^*(t) - \bar{F}(t)\}/\bar{F}(t) \\ = - \int_0^t n^{1/2} \frac{\bar{F}_{n0}^*(u-)m(u; \theta_0)I(R(u) > 0)}{\bar{F}(u)R(u)} dM(u) \\ + n^{1/2}I(T_n < t)\bar{F}_{n0}^*(T_n)\{\bar{F}(t)\}^{-1}\{\bar{F}(T_n)\}^{-1}\{\bar{F}(T_n) - \bar{F}(t)\} \\ - \bar{F}_{n0}^*(t)\{\bar{F}(t)\}^{-1}U_n'(t; \bar{\theta})\{n^{1/2}(\bar{\theta} - \theta_0)\},$$

where

$$U_n(t; \theta) = \int_0^t \{\bar{F}_{n0}^*(u)\}^{-1} \bar{F}_n^*(u-) \frac{\partial m(u; \theta)}{\partial \theta} d\Lambda_n(u).$$

Since

$$\sup_{0 \leq u \leq t} \left| \left[n^{1/2} \frac{\bar{F}_{n0}^*(u-)m(u; \theta_0)I(R(u) > 0)}{\bar{F}(u)R(u)} \right]^2 R(u) \right. \\ \left. - \frac{\{m(u; \theta_0)\}^2}{\bar{H}(u-)\{1 - m(u; \theta_0)\Delta\Lambda(u)\}^2} \right| \xrightarrow{P} 0,$$

as $n \rightarrow \infty$, the asymptotic distribution of the first term on the right-hand side of (A.5) is normal with mean zero and variance given by the second term on the right-hand side of (3.6). Since $U_n(t; \bar{\theta})$ converges to $U(t; \theta_0)$ in probability, the asymptotic distribution of the third term on the right-hand side of (A.5) is normal with mean zero and variance given by the first term on the right-hand side of (3.6). The second term on the right-hand side of (A.5) is negligible because

$$0 \leq n^{1/2}I(T_n < t)\bar{F}_{n0}^*(T_n)\{\bar{F}(t)\}^{-1}\{\bar{F}(T_n)\}^{-1}\{\bar{F}(T_n) - \bar{F}(t)\} \\ \leq n^{1/2}I(T_n < t)\{\bar{F}(t)\}^{-2}\{1 - \bar{F}(t)\} \xrightarrow{P} 0, \quad \text{as } n \rightarrow \infty.$$

Denote the first term on the right-hand side of (A.5) by $B_n(t)$. Then we have

$$\langle S_n(\theta_0, \cdot), B_n \rangle(t) = 0,$$

and hence the score vector process $S_n(\theta_0, t)$ and the process $B_n(t)$ are orthogonal. Thus we can see that the first term and the third term are asymptotically independent. $\square$

Acknowledgements

We would like to extend our thanks to the referees, who made a number of helpful comments.

REFERENCES

- [1] Aalen O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701-726.
- [2] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory*, Ed. B. N. Petrov and F. Csaki, 267-81. Budapest, Akademiai Kiado. (Reproduced (1992) in *Breakthroughs in*

- Statistics*, **1**, (S. Kotz and N. L. Johnson, eds.), 610-624. New York, Springer-verlag.)
- [3] Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *Ann. Statist.*, **2**, 437-453.
 - [4] Efron, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkely Symp. Math. Statist. Prob.*, **4**, 831-853.
 - [5] Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
 - [6] Freireich, E. J., Gehan, E., Frei, E., Schroeder, L. R., Wolman, I. J., Anbari, R., Burgert, E. O., Mills, S. D., Pinkel, D., Selawry, O. S., Moon, J. H., Gendel, B. R., Spurr, C. L., Storrs, R. Haurani, F., Hoogstraten, B. and Lee, S. (1963). The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy. *Blood*, **21**, 699-716.
 - [7] Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily single-censored sample. *Biometrika*, **52**, 203-223.
 - [8] Gill, R. D. (1980). *Censoring and Stochastic Integrals*. Math. Centre Tracts, **124**, Mathematisch Centrum, Amsterdam.
 - [9] Gill, R. D. (1983). Large sample behavior of the product limit estimator on the whole line. *Ann. Statist.*, **11**, 49-58.
 - [10] Hurvich, C. M. and Tsai, C-L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76**, 297-307.
 - [11] Johansen, S. (1978). The product limit estimator as maximum likelihood estimator. *Scand. J. Statist.*, **5** 195-199.
 - [12] Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.
 - [13] Kaplan, E. L. and Meier, P. (1958). Non-parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, 457-481.
 - [14] Klein, J. P. (1991). Small-sample moments of some estimators of the variance of the Kaplan-Meier and Nelson-Aalen Estimators. *Scand. J. Statist.*, **18**, 333-340.
 - [15] Lenglar, E. (1977). Relation de domination entre deux processus. *Ann. Inst. Henri Poincaré*, **13**, 171-9.
 - [16] Miller, R. G. (1983). What price Kaplan-Meier?. *Biometrics*, **39**, 1077-81.
 - [17] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, **14**, 945-965.
 - [18] Reid, N. (1981). Influence functions for censored data. *Ann. Statist.*, **9**, 78-92.
 - [19] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461-464.
 - [20] Stute, W. (1994). The bias of Kaplan-Meier integrals. *Scand. J. Statist.*, **21**, 475-484.
 - [21] Stute, W. and Wang, J. L. (1993). The strong law under random censorship. *Ann. Statist.*, **21**, 1591-1607.
 - [22] Wang, J. (1987). A note on the uniform consistency of the Kaplan-Meier estimator. *Ann. Statist.*, **15**, 1313-1316.