Journal of the Korean
OR/MS Society
Vol. 14 No. 1 June 1989

—Research Paper—

Generalized Maximum Likelihood Estimation in a Multistate Stochastic Model

Sungchil Yeo*

ABSTRACT

Multistate survival data with censoring often arise in biomedical experiments. In particular, a four-state space is used for cancer clinical trials. In a four-state space, each patient may either respond to a given treatment and then relapse or may progress without responding. In this four-state space, a model which combines the Markov and semi-Markov models is proposed. In this combined model, the generalized maximum likelihood estimators of the Markov and semi-Markov hazard functions are derived. These estimators are illustrated for the data collected in a study of treatments for advanced breast cancer.

1. Introduction

In clinical trials and other medical studies, one may often be concerned with the evaluation of two or more successive event times and their relationships to one another. For example, in cancer clinical trials, people commonly study the time that a patient reaches a specific illness state and the amount of time that the patient stayes in that state. The statistical analysis of such two or more time-dependent events is called multistate survival analysis.

Over the last few decades, various kinds of stochastic models have been proposed for analyzing multistate survival data by many people. Lagakos (1976, 1977) applied a homogeneous Markov model to analyze survival data in the presence of auxiliary information. Temkin (1978) considered a non-homogeneous Markov model and proposed the probability of being in response function (PBRF) as a summary description for assessing the response to a treatment in cancer clinical trials. Begg and Larson (1982) examined the properties of the PBRF based on a homogeneous Markov model and demonstrated that the PBRF is a fairly complete description of the effects of treatments. Aalen and Johansen (1978) suggested a product limit estimator and studied its properties for the transition probabilities of a more general non-homogeneous Markov model with censored observations. Fleming (1978 a, b) also studied the same product limit estimator based on the non-homogeneous Markov process as Aalen and Johansen (1978) studied, but in the absence of censoring. On the other hand, to analyze data obtained from a certain clinical trials, Weiss and Zelen (1965) proposed a semi-Markov model. Lagakos et al. (1978) suggested a nonparametric likelihood method for the analysis of partially censored data

* Department of Applied Statistics, Kon-Kuk University, Seoul, Korea

based on a semi-Markov process model. Nonparametric estimation for the semi-Markov process was further developed in a more general setting by Gill (1980). Voelkel and Crowley (1984) applied a counting process approach to some hierarchical semi-Markov processes and obtained some useful asymptotic results for the PBRF. In this paper, we propose a combined model which is an extension of the Markov and semi-Markov models. However, in order to simplify ideas, we define our model only in a four-state space used by Temkin (1978) for clinical trials.

In a four-state space, all patients who start in an initial state, 0, may be assigned at random to receive a drug treatment. Upon receiving this treatment, a patient's condition may either deteriorate without showing any improvement (progressive state, 3) or may initially improve (response state, 1) and afterward deteriorate (relapse state, 2). We assume that for any patient, one of these transitions would occur with probability one, but that patient may be censored before the transition occurs. We also assume that no patients can be in the response state after either progression or relapse. Figure 1 exhibits this four-state space. In Figure 1, R, S, and T represent the patient's time to the progressive, response, and relapse states, respectively. At time t, the hazard rate is defined as follows :

$$\lambda_S(t) = \lim_{h \to 0^+} \frac{1}{h} P(t \leq S < t+h \quad S \geq t) \tag{1.1}$$

Thus $\lambda_S(t)$ represents the instantaneous rate of transition to the response state at time t, conditional upon the patient's survival until time t. The hazard rate $\lambda_R(t)$ can be described similarly. The conditional hazard rate is defined as follows :

$$\lambda(t \mid s) = \lim_{h \to 0^+} \frac{1}{h} P(t \leq T < t+h \mid T \geq t, \quad S=s), \quad s \leq t \tag{1.2}$$

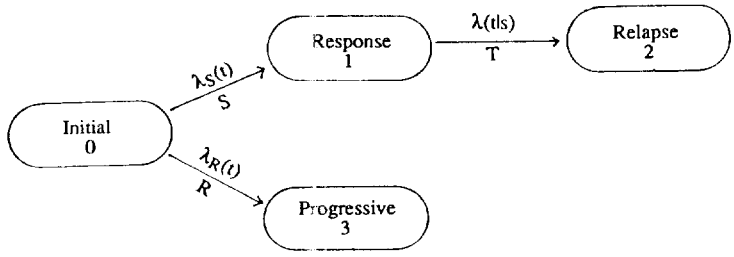

Figure 1. Four-state space

Thus $\lambda(t \mid s)$ specifies the instantaneous rate of transition at time t to the relapse state, conditional on having entered the response state at time s and survival until time t. In this four-state space, homogeneous Markov, non-homogeneous Markov, and semi-Markov models, as well as our "combined model", can now be described in terms of hazard rates, completely analogous

to the transition probabilities in a more general stochastic process. If $\lambda(t \mid s)$, $\lambda_S(t)$, and $\lambda_R(t)$ are constants, then the model is said to be a homogeneous Markov model. If $\lambda(t \mid s) = \lambda_M(t)$, a function of t only, then the model is said to be a non-homogeneous Markov model. If $\lambda(t \mid s) = \lambda_{SM}(t-s)$, a function of t−s, then the model is said to be a semi-Markov model. Now the combined model proposed in this paper is defined as follows :

$$\lambda(t \mid s) = \lambda_M(t) + \lambda_{SM}(t-s), \quad s < t \tag{1.3}$$

The model (1.3) can be motivated if the patient's condition, after reaching the response state, depends on both "absolute time", the time since the start of the study, and "duration time", the time elapsed since the entry into the present state. Figure 2 illustrates the homogeneous Markov, non-homogeneous Markov, and semi-Markov models as well as the combined model in four-state space.

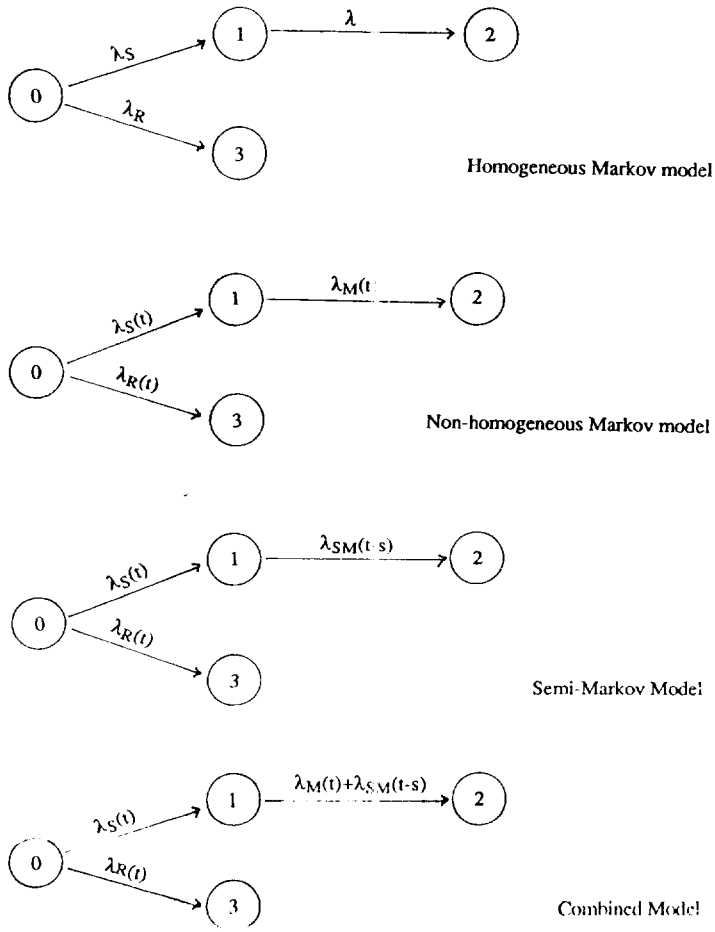

Figure 2. Four models in four-state space

2. Generalized Maximum Likelihood Estimation for the combined model

In this section, we derive the generalized maximum likelihood estimators (GMLE) of the Markov and semi-Markov hazard functions in the combined model. Before we obtain the GMLE's, we review the definition of the GMLE and give several example.

In the usual parametric models, the observed data x has a probability measure such that

$$\frac{dP_\theta(x)}{d\mu(x)} = f_\theta(x), \tag{2.1}$$

where, μ is a dominating measure for the family of probality measure $\mathscr{P}_\theta$. Then the maximum likelihood estimator (MLE) of θ is defined as the value which maximizes (2.1). In nonparametric models, the observed data x has a probability measure which depends on the unknown distribution function F. For a nondominated family of probability measures $\mathscr{P}$, the usual likelihood is not defined. Thus, we need a more general definition of maximum likelihood estimation.

Kiefer and Wolfowitz (1956) suggested that for a nondominated family of probability measures $\mathscr{P}$, one can define a generalized maximum likelihood estimator as follows. For P_1 and P_2 in $\mathscr{P}$, define

$$f(x; P_1, P_2) = \frac{dP_1}{d(P_1+P_2)}(x), \tag{2.2}$$

the Radon-Nikodym derivative of P_1 with respect to P_1+P_2. If x represents the observed data, then $\hat{P}$ is a GMLE if

$$f(x; \hat{P}, P) \geq f(x; P, \hat{P}) \qquad \text{for all } P \in \mathscr{P} \tag{2.3}$$

A distribution function $\hat{F}$ is said to be a GMLE if $\hat{F}$ induced by the probability measure $\hat{P}$. The definition (2.3) of the GMLE reduces to the usual definition (2.1) of the MLE if $\mathscr{P}$ is dominated by the σ-finite measure μ.

Johansen (1978) pointed out that if $\hat{P}$ gives positive probability to the observed data x, then $f(x; P, \hat{P}) = 0$ unless P also gives positive probability to x. Thus, in order to check (2.3), it is enough to check it for those $P \in \mathscr{P}$ with $P\{x\} > 0$ provided that the family of such P is nonempty. In this case, (2.3) reduces to

$$\hat{P}\{x\} \geq P\{x\}. \tag{2.4}$$

There are several examples for the GMLE's. The empirical distribution function $F_n(t) = (\text{numbers of } X_i \leq t)/n$ is the GMLE of F in the class of all distribution functions. This estimator is unbiased, consistent, and asymptotically normal. Johansen (1978) showed that the product limit estimator suggested by Kaplan and Meier (1958) is the GMLE of $\bar{F}$ in the class of all survival functions with censored observations. Peterson (1977) showed that the product limit estimator is strongly consistent. Breslow and Crowly (1974) established the asymptotic normality of this estimator. In nonparametric inference for multistate survival data, Voelkel (1980)

showed that the GMLE's of the cumulative Markov and semi-Markov hazard functions in the purely Markov and purely semi-Markov model, respectively, are consistent and asymptotically normal, although he did not use the generalized maximum likelihood estimation method.

We now derive the GMLE's of the cumulative Markov and semi-Markov hazard functions in the combined model. In section 1, the combined model (1.3) was described for the continuous case. However, if we define the combined model only for the continuous case, then the GMLE's of $\bar{F}_M(t)$ and $\bar{F}_{SM}(u)$ do not exist. Thus, in order to obtain the GMLE's of $\bar{F}_M(t)$ and $\bar{F}_{SM}(u)$, we rewrite the combined model in terms of survival functions so that the model includes both continuous and discrete cases. Define the combined model as

$$\bar{\mathbf{F}}(t \mid s) = \frac{\bar{F}_M(t)}{\bar{F}_M(s)} \bar{F}_{SM}(t-s), \quad 0 \leqq s \leqq t < \infty, \tag{2.5}$$

where $\bar{F}_M(t)$ and $\bar{F}_{SM}(u)$ are the survival functions corresponding to the Markov and semi-Markov hazard functions $\lambda_M(t)$ and $\lambda_{SM}(u)$, respectively. If the underlying distributions of then model (2.5) are continuous, then the model (2.5) reduces to the model (1.3). If the underlying distributions of the model (2.5) are discrete, then the model (2.5) can be written in terms of hazard functions by

$$1-\lambda(t \mid s) = (1-\lambda_M(t))\,(1-\lambda_{SM}(t-s)), \quad 0 \leqq s \leqq t < \infty. \tag{2.6}$$

We now proceed to find the nonparametric likelihood for four-state survival data. Suppose there are n patients in the initial state at time 0. Let r_i, s_i, and t_i be the i^{th} patient's times to the progressive, response, and relapse states, respectively. As usual, we cannot observe r_i, s_i, and t_i completely. To include censoring, we assume that the censoring times are univariate, that is, the censoring times for the periods from one state to next state are identical. Let c_i be the censoring time of the i^{th} patient. Then the observations on n patients consist of the pairs $(w_i, x_i, y_i, \delta_{1i}, \delta_{2i}, \delta_{3i})$, $i=1,2,\cdots,n$, where $w_i = r_i \wedge c_i$, $x_i = s_i \wedge c_i$, $y_i = t_i \wedge c_i$, $\delta_{1i} = [s_i \leqq c_i]$, $\delta_{2i} = [t_i \leqq c_i]$, $\delta_{3i} = [r_i \leqq c_i]$, and where $x \wedge y$ denotes $\min(x, y)$, and $[A]$ denotes the indicator function of the set A. We also assume that the observed vector of transition times (r_i, s_i, t_i) and the censoring time c_i are independent for each i. Then, the nonparametric likelihood for four-state survival data is proportional to

$$\begin{aligned}
\boldsymbol{L} &= \prod_{i=1}^{n} \{\bar{F}_S(x_i+0)\ \bar{F}_R(w_i+0)\}^{(1-\delta_{1i})(1-\delta_{3i})} \{\bar{F}_S(x_i+0)\ d\bar{F}_R(w_i)\}^{\delta_{3i}} \\
&\quad \times \{\bar{F}(y_i+0 \mid x_i)\ \bar{F}_R(w_i+0)\ d\bar{F}_S(x_i)\}^{\delta_{1i}(1-\delta_{2i})} \{\bar{F}_R(w_i+0)\ d\bar{F}_S(x_i)\ d\bar{F}(y_i \mid x_i)\}^{\delta_{1i}\delta_{2i}} \\
&= \prod_{i=1}^{n} \{\bar{F}_S(x_i+0)\}^{(1-\delta_{1i})}\ \{d\bar{F}_S(x_i)\}^{\delta_{1i}} \\
&\quad \times \prod_{i=1}^{n} \{\bar{F}_R(w_i+0)\}^{(1-\delta_{3i})}\ \{d\bar{F}_R(w_i)\}^{\delta_{3i}} \\
&\quad \times \prod_{i=1}^{n} \{\bar{F}(y_i+0 \mid x_i)\}^{\delta_{1i}(1-\delta_{2i})}\ \{d\bar{F}(y_i \mid x_i)\}^{\delta_{1i}\delta_{2i}},
\end{aligned} \tag{2.7}$$

where $\bar{F}(t+0)=\lim_{h\to 0^+}\bar{F}(t+h)$, and where $d\bar{F}(t)$ equals $-\bar{F}'(t)$ if $\bar{F}(t)$ is absolutely continuous at t and $\bar{F}(t)-\bar{F}(t+0)$ if $\bar{F}(t)$ has a jump at t.

Since each product term in (2.7) involves only one of the functions $\bar{F}_S(\cdot)$, $\bar{F}_R(\cdot)$, and $\bar{F}(\cdot|\cdot)$, L is maximized by maximizing each product term in (2.7). However, since we are mainly interested in finding estimators of $\Lambda_M(\cdot)$, and $\Lambda_{SM}(\cdot)$, we only need to maximize the part of the likelihood for the period from response to relapse, which is proportional to

$$L_R=\prod_i^n \{\bar{F}(y_i+0\mid x_i)\}^{\delta_{1i}(1-\delta_{2i})}\{d\bar{F}(y_i\mid x_i)\}^{\delta_{1i}\delta_{2i}}. \tag{2.8}$$

Based on (2.8), the GMLE's of $\Lambda_M(\cdot)$ and $\Lambda_{SM}(\cdot)$ are provided by the following theorem.

Theorem 2.1. Under the combined model (2.5), if at least one patient is observed to relapse, then the GMLE's of $\Lambda_M(\cdot)$ and $\Lambda_{SM}(\cdot)$ are as follows:

$$\hat{\Lambda}_M(t)=\sum_i^n \hat{\lambda}_i^M \delta_{2i}[y_i<t], \tag{2.9}$$

and

$$\hat{\Lambda}_{SM}(u)=\sum_i^n \hat{\lambda}_i^{SM} \delta_{2i}[z_i<u], \tag{2.10}$$

where if $A_i\neq B_i$, then

$$\hat{\lambda}_i^M=\frac{\delta_{1i}}{A_i}[A_i<B_i], \tag{2.11}$$

$$\hat{\lambda}_i^{SM}=\frac{\delta_{1i}}{B_i}[A_i<B_i], \tag{2.12}$$

and if $A_i=B_i$, $\hat{\lambda}_i^M$ and $\hat{\lambda}_i^{SM}$ may be any values between 0 and 1 such that

$$(1-\hat{\lambda}_i^M)(1-\hat{\lambda}_i^{SM})=1-\frac{\delta_{1i}}{A_i}, \tag{2.13}$$

and where,

$$A_i=\sum_j^n \delta_{1j}[x_i\leq y_j\leq y_i], \tag{2.14}$$

$$B_i=\sum_j^n \delta_{1j}[z_j\leq z_i],\quad z_i=y_i-x_i. \tag{2.15}$$

If no patients are observed to relapse, then the GMLE's are not defined.

In (2.14) and (2.15), A_i is the number of patients at risk in state 1 just before time y_i, and B_i is the number of patients whose duration time in state 1 is at least z_i. In Theorem 2.1, when

$A_i=B_i$, the GMLE's of $\Lambda_M(t)$ and $\Lambda_{SM}(u)$ are not unique. In this situation, two convenient choices for $\hat{\lambda}_i^M$ and $\hat{\lambda}_i^{SM}$ are $\hat{\lambda}_i^M=\delta_{1i}/A_i$, $\hat{\lambda}_i^{SM}=0$ and $\hat{\lambda}_i^M=0$, $\hat{\lambda}_i^{SM}=\delta_{1i}/A_i$. If we take $\hat{\lambda}_i^M=\delta_{1i}/A_i$, $\hat{\lambda}_i^{SM}=0$ when $A_i=B_i$, the GMLE's of $\Lambda_M(t)$ and $\Lambda_{SM}(u)$ are as follows :

$$\hat{\Lambda}_M(t)=\sum_{i=1}^{n}\frac{\delta_{1i}\,\delta_{2i}}{A_i}[A_i\leq B_i]\,[Y_i<t] \tag{2.16}$$

and

$$\hat{\Lambda}_{SM}(u)=\sum_{i=1}^{n}\frac{\delta_{1i}\,\delta_{2i}}{B_i}[A_i>B_i]\,[Z_i<u]\,. \tag{2.17}$$

For the special cases, under the purely Markov and purely semi-Markov models, the corresponding GMLE's of $\Lambda_M(t)$ and $\Lambda_{SM}(u)$ are

$$\hat{\Lambda}_M(t)=\sum_{i=1}^{n}\frac{\delta_{1i}\,\delta_{2i}}{A_i}[Y_i<t] \tag{2.18}$$

and

$$\hat{\Lambda}_{SM}(u)=\sum_{i=1}^{n}\frac{\delta_{1i}\,\delta_{2i}}{B_i}[Z_i<u]\,, \tag{2.19}$$

respectively.

The following lemma will be needed for the proof of Theorem 2.1. Its proof is given in Appendix A.

Lemma 2.1. Let A, B, and δ be constants such that $\min(A, B)\geq 1$ and δ is either 0 or 1. Define

$$\begin{aligned} g(\alpha,\beta)=&\ \delta\ln\left(\frac{1}{(1-\alpha)(1-\beta)}-1\right)\\ &+A\ln(1-\alpha)+B\ln(1-\beta)\,, \end{aligned} \tag{2.20}$$

where α and β are unknown parameters in $[0, 1)$. Then the maximum of $g(\alpha, \beta)$ is attained at $(\hat{\alpha}, \hat{\beta})$, where if $A\neq B$, then

$$\hat{\alpha}=\frac{\delta}{A}[A<B]\,,\quad \hat{\beta}=\frac{\delta}{B}[A>B]\,,$$

and if $A=B$, then $\hat{\alpha}$ and $\hat{\beta}$ may be any values in $[0, 1)$ such that

$$(1-\hat{\alpha})(1-\hat{\beta})=1-\frac{\delta}{A}\,.$$

We now prove Theorem 2.1.

Proof of Theorem 2.1 : Under the combined model (2.5), (2.8) can be written as

$$L_R=\prod_{i=1}^{n}\{\frac{\bar{F}_M(y_i+0)}{\bar{F}_M(x_i)}\ \bar{F}_{SM}(y_i-x_i+0)\}^{\delta_{1i}(1-\delta_{2i})}$$

$$\times\{d\,(\frac{\bar{F}_M(y_i)}{\bar{F}_M(x_i)}\ \bar{F}_{SM}(y_i-x_i))\}^{\delta_{1i}\,\delta_{2i}} \tag{2.21}$$

In (2.21), as Johansen (1978) pointed out, in order to find the GMLE's of $\bar{F}_M(t)$ and $\bar{F}_{SM}(u)$, we only need to consider $\bar{F}_M(t)$ and $\bar{F}_{SM}(u)$ with jumps only at the observed transition times. Let

$$\bar{F}_M(t)=\prod_{\substack{y_j<t\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M) \tag{2.22}$$

and

$$\bar{F}_{SM}(u)=\prod_{\substack{z_j<u\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})\,, \tag{2.23}$$

where $\lambda_i^M=\lambda_M(y_i)$ and $\lambda_i^{SM}=\lambda_{SM}(z_i)$. Then (2.21) implies that

$$L_R=\prod_{i=1}^{n}\{\frac{\bar{F}_M(y_i+0)}{\bar{F}_M(x_i)}\ \bar{F}_{SM}(y_i-x_i+0)\}^{\delta_{1i}(1-\delta_{2i})}$$

$$\times\{\frac{\bar{F}_M(y_i)}{\bar{F}_M(x_i)}\ \bar{F}_{SM}(y_i-x_i)-\frac{\bar{F}_M(y_i-0)}{\bar{F}_M(x_i)}\ \bar{F}_{SM}(y_i-x_i+0)\}^{\delta_{1i}\,\delta_{2i}}$$

$$=\prod_{i=1}^{n}\{\prod_{\substack{x_j<y_j\le y_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M)\prod_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})\}^{\delta_{1i}(1-\delta_{2i})}$$

$$\times\{\prod_{\substack{x_j\le y_j\le y_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M)\prod_{\substack{z_j<z_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})-\prod_{\substack{x_j\le y_j\le y_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M)\prod_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})\}^{\delta_{1i}(1-\delta_{2i})}$$

$$=\prod_{i=1}^{n}\{\prod_{\substack{x_j\le y_j\le y_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M)\prod_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})\}^{\delta_{1i}(1-\delta_{2i})}$$

$$\times\{\prod_{\substack{x_j\le y_j\le y_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M)\prod_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})\,(\frac{1}{(1-\lambda_i^M)(1-\lambda_i^{SM})}-1)\}^{\delta_{1i}\,\delta_{2i}} \tag{2.24}$$

If we take the logarithm of L_R in (2.24), then (2.24) implies that

$$ln\,L_R=\sum_{i=1}^{n}\delta_{1i}(1-\delta_{2i})\,ln\,\{\prod_{\substack{x_j\le y_j\le y_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M)\prod_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})\}$$

$$+\sum_{i=1}^{n}\delta_{1i}\delta_{2i}\ln\{\prod_{\substack{x_i\le y_j\le y_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M)\}\{\prod_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})\left(\frac{1}{(1-\lambda_j^M)(1-\lambda_j^{SM})}-1\right)\}$$

$$=\sum_{i=1}^{n}\delta_{1i}(1-\delta_{2i})\{\sum_{\substack{x_i\le y_j\le y_i\\ \&\ \delta_{2j}=1}}\ln(1-\lambda_j^M)\sum_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}\ln(1-\lambda_j^{SM})\}$$

$$+\sum_{i=1}^{n}\delta_{1i}\delta_{2i}\{\sum_{x_i\le y_j\le y_i}\ln(1-\lambda_j^M)+\sum_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}\ln(1-\lambda_j^{SM})\}$$

$$+\sum_{i=1}^{n}\delta_{1i}\delta_{2i}\ln\left(\frac{1}{(1-\lambda_i^M)(1-\lambda_i^{SM})}-1\right)$$

$$=\sum_{i=1}^{n}\delta_{2i}\ln(1-\lambda_i^M)\sum_{i=1}^{n}\delta_{1i}[x_i\le y_i\le y_i]$$

$$+\sum_{i=1}^{n}\delta_{2i}\ln(1-\lambda_i^{SM})\sum_{i=1}^{n}\delta_{1i}[z_i\le z_i]$$

$$+\sum_{i=1}^{n}\delta_{1i}\delta_{2i}\ln\left(\frac{1}{(1-\lambda_i^M)(1-\lambda_i^{SM})}-1\right) \tag{2.25}$$

From (2.14) and (2.15), (2.25) can be rewritten as

$$\ln L_R=\sum_{i=1}^{n}\{\delta_{1i}\delta_{2i}\ln\left(\frac{1}{(1-\lambda_i^M)(1-\lambda_i^{SM})}-1\right)+\delta_{2i}A_i\ln(1-\lambda_i^M)+\delta_{2i}B_i(1-\lambda_i^{SM})\} \tag{2.26}$$

Define, for each j,

$$g_j(\lambda_j^M,\lambda_j^{SM})=\delta_{1j}\ln\left(\frac{1}{(1-\lambda_j^M)(1-\lambda_j^{SM})}-1\right)+A_j\ln(1-\lambda_j^M)+B_j(1-\lambda_j^{SM})$$

Then (2.26) reduces to

$$\ln L_R=\sum_{i=1}^{n}\delta_{2i}g_i(\lambda_i^M,\lambda_i^{SM}) \tag{2.27}$$

Since we are interested in finding pairs $(\lambda_i^M,\lambda_i^{SM})$ which maximize (2.27) and since the $(\lambda_i^M,\lambda_i^{SM})$'s have no relationship with each other, it suffices to fine λ_i^M and λ_i^{SM} which maximizes $g_i(\lambda_i^M,\lambda_i^{SM})$ is attained at $(\lambda_i^M,\lambda_i^{SM})$ which is given by (2.11), (2.12), and (2.13). Thus, the GMLE's of $\bar{F}_M(t)$, and $\bar{F}_{SM}(u)$, are

$$\hat{\bar{F}}_M(t)=\prod_{\substack{y_j<t\\ \&\ \delta_{2j}=1}}(1-\hat{\lambda}_j^M)$$

and

$$\hat{F}_{SM}(u) = \prod_{\substack{z_i < u \\ \& \ \delta_i = 1}} (1 - \hat{\lambda}_i^{SM})$$

Therefore, the corresponding GMLE's of $\Lambda_M(t)$ and $\Lambda_{SM}(u)$ are given by (2.9) and (2.10).

3. Example

In this section, we illustrate the GMLE's for certain actual data (Hsieh (1980)). In a clinical trial performed at the Wisconsin Clinical Cancer Center, 135 patients with advanced breast cancer were treated. Of the 135 patients, 55 were treated with Dibromodulcitol and Adriamycin (DA), and 80 were treated with Dibromodulcitol, Adriamycin, and Tamoxifen (DAT). In this study, the ECOG (Eastern Cooperative Oncology Group) criteria were used to evaluate each patient's response. Among the 55 DA patients, 18 relapsed after having responded, 2 responded without relapse, and 35 progressed. Among the 80 DAT patients, 24 relapsed after having res-

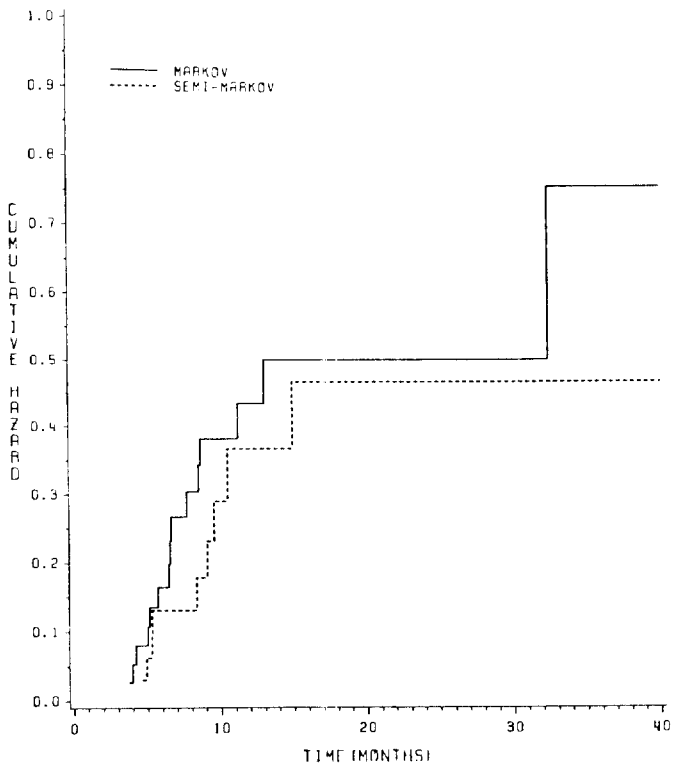

Figure 3
Gmles of the cumulative markov and semimarkov hazard functions in the combined model.

$$+\sum_{i=1}^{n}\delta_{1i}\delta_{2i}\ln\{\prod_{\substack{x_i\le y_j\le y_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^M)\}\{\prod_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}(1-\lambda_j^{SM})\left(\frac{1}{(1-\lambda_j^M)(1-\lambda_j^{SM})}-1\right)\}$$

$$=\sum_{i=1}^{n}\delta_{1i}(1-\delta_{2i})\{\sum_{\substack{x_i\le y_j\le y_i\\ \&\ \delta_{2j}=1}}\ln(1-\lambda_j^M)\sum_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}\ln(1-\lambda_j^{SM})\}$$

$$+\sum_{i=1}^{n}\delta_{1i}\delta_{2i}\{\sum_{x_i\le y_j\le y_i}\ln(1-\lambda_j^M)+\sum_{\substack{z_j\le z_i\\ \&\ \delta_{2j}=1}}\ln(1-\lambda_j^{SM})\}$$

$$+\sum_{i=1}^{n}\delta_{1i}\delta_{2i}\ln\left(\frac{1}{(1-\lambda_i^M)(1-\lambda_i^{SM})}-1\right)$$

$$=\sum_{i=1}^{n}\delta_{2i}\ln(1-\lambda_i^M)\sum_{i=1}^{n}\delta_{1i}[x_i\le y_i\le y_i]$$

$$+\sum_{i=1}^{n}\delta_{2i}\ln(1-\lambda_i^{SM})\sum_{i=1}^{n}\delta_{1i}[z_i\le z_i]$$

$$+\sum_{i=1}^{n}\delta_{1i}\delta_{2i}\ln\left(\frac{1}{(1-\lambda_i^M)(1-\lambda_i^{SM})}-1\right) \tag{2.25}$$

From (2.14) and (2.15), (2.25) can be rewritten as

$$\ln L_R=\sum_{i=1}^{n}\{\delta_{1i}\delta_{2i}\ln\left(\frac{1}{(1-\lambda_i^M)(1-\lambda_i^{SM})}-1\right)+\delta_{2i}A_i\ln(1-\lambda_i^M)+\delta_{2i}B_i(1-\lambda_i^{SM})\} \tag{2.26}$$

Define, for each j,

$$g_j(\lambda_j^M,\lambda_j^{SM})=\delta_{1j}\ln\left(\frac{1}{(1-\lambda_j^M)(1-\lambda_j^{SM})}-1\right)+A_j\ln(1-\lambda_j^M)+B_j(1-\lambda_j^{SM})$$

Then (2.26) reduces to

$$\ln L_R=\sum_{i=1}^{n}\delta_{2i}g_i(\lambda_i^M,\lambda_i^{SM}) \tag{2.27}$$

Since we are interested in finding pairs $(\lambda_i^M,\lambda_i^{SM})$ which maximize (2.27) and since the $(\lambda_i^M,\lambda_i^{SM})$'s have no relationship with each other, it suffices to fine λ_i^M and λ_i^{SM} which maximizes $g_i(\lambda_i^M,\lambda_i^{SM})$ is attained at $(\lambda_i^M,\lambda_i^{SM})$ which is given by (2.11), (2.12), and (2.13). Thus, the GMLE's of $\bar{F}_M(t)$, and $\bar{F}_{SM}(u)$, are

$$\hat{\bar{F}}_M(t)=\prod_{\substack{y_j<t\\ \&\ \delta_{2j}=1}}(1-\hat{\lambda}_j^M)$$

Now if we differentiate $h^*(u)$ with respect to u, then

$$\frac{dh^*(u)}{du} = \frac{\delta}{u-1} + \frac{A-\delta}{u}. \qquad (A.5)$$

Solving $dh^*(u)/du=0$, we obtain $u=1-\delta/A$. Since $d^2h^*(u)/du^2<0$ at $u=1-\delta/A$, the maximum of $h^*(u)$ is attained at $u=1-\delta/A$. Therefore the maximum of $h(u, w)$ is attained at $u=1-\delta/A$ and $w=1$. In this case, the maximum of $g(\alpha, \beta)$ is attained at $\hat{\alpha}=\delta/A$ and $\hat{\beta}=0$.

Case 2 : A>B.

By the same arguments as Case 1, the maximum of $g(\alpha, \beta)$ is attained at $\hat{\alpha}=0$ and $\hat{\beta}=\delta/B$.

Case 3 : A=B.

From (A.1), $h(u, w)$ reduces to

$$h^*(u) = \delta \ln(1-u) + (A-\delta)\ln u. \qquad (A.6)$$

Thus, from the results of Case 1, $h^*(u)$ has the maximum value at $u=1-\delta/A$. In this case, the maximum of $g(\alpha, \beta)$ is attained at any pairs $(\hat{\alpha}, \hat{\beta})$ in $[0, 1)$ such that $(1-\hat{\alpha})(1-\hat{\beta})=1-\delta/A$.

Appendix B

Table 1. Advanced Breast Cancer Data *

Patient's Number	S	T	U	Censoring Code
1	199	458	259	0
2	84	399	315	1
3	84	196	112	1
4	31	119	88	1
5	−1	55	−1**	1
6	−1	98	−1	1
7	28	167	139	1
8	28	231	203	1
9	−1	84	−1	1
10	21	180	159	1
11	28	315	287	1
12	59	171	112	1
13	87	150	63	1
14	28	343	315	0
15	259	555	296	0
16	31	191	160	1
17	−1	28	−1	1
18	−1	141	−1	1
19	28	941	913	0
20	−1	38	−1	1

Patient's Number	S	T	U	Censoring Code
21	−1	42	−1	1
22	−1	87	−1	1
23	−1	28	−1	1
24	56	504	448	1
25	59	199	140	1
26	−1	114	−1	1
27	28	175	147	1
28	143	255	112	1
29	−1	84	−1	1
30	−1	476	−1	0
31	−1	99	−1	1
32	80	283	203	0
33	37	326	289	0
34	112	337	225	0
35	91	364	273	0
36	−1	71	−1	1
37	−1	47	−1	1
38	28	126	98	1
39	56	970	914	1
40	31	143	112	0
41	−1	85	−1	1
42	56	462	406	0
43	84	154	70	1
44	38	1290	1252	0
45	93	211	118	0
46	103	390	287	1
47	−1	68	−1	1
48	−1	37	−1	1
49	−1	98	−1	1
50	112	1127	1015	0
51	−1	182	−1	1
52	60	112	52	1
53	8	259	251	1
54	−1	28	−1	1
55	−1	59	−1	1
56	−1	63	−1	1
57	−1	125	−1	1
58	119	336	217	1
59	58	540	482	0
60	28	191	163	0
61	−1	168	−1	1
62	−1	60	−1	1
63	81	193	112	1
64	175	627	452	0
65	−1	80	−1	1
66	27	582	555	0
67	−1	32	−1	1
68	7	279	272	1
69	−1	56	−1	1
70	86	597	511	0
71	−1	84	−1	1

72	−1	22	−1	1
73	111	334	223	0
74	−1	122	−1	1
75	−1	35	−1	1
76	56	259	203	1
77	−1	28	−1	1
78	−1	111	−1	1
79	140	1134	994	0
80	−1	63	−1	1

* S=time to reponse (in days) ;
T=time to progression (relapse) or censoring (in days) ;
U=duration of response (in days) ;
Censoring Code : 1=progression or relapse, 0=censoring.

** −1=no response.

REFERENCES

1. Aalen, O. and Johansen, S. (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scand. J. Statist.* 5, 141-150.
2. Begg, C. B. and Larson, M. (1982). A study of the use of the probability-of-being-in-response function as a summary of tumor response data. ***Biometrics***, 38, 59-66.
3. Breslow, N. and Crowley, J. (1974). A large-sample study of the life table and product limit estimates under random censorship. *Ann. Math. Statist.* 2, 437-453.
4. Fleming, T. R. (1978a). Nonparametric estimation for nonhomogeneous Markov processes in the problem of competing risks. *Ann. Statist.* 6, 1057-1070.
5. Fleming, T. R. (1978b). Asymptotic destribution results in competing risks estimation. *Ann. Statist.* 6, 1071-1079.
6. Gill, R. D. (1980). Nonparametric estimation based on censored observations of a Markov renewal process. *Z. Wahrsch. verw. Gebiete* 53, 97-116.
7. Hsieh, F-Y. (1980). Multistate Survival Anslysis. Ph. D., thesis Department of Statistics, University of Wisconsin-Madison.
8. Johansen, S. (1978). The product limit estimator as maximum likelihood estimator. *Scand. J. Statist.* 5, 195-199.
9. Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.
10. Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53, 457-481.
11. Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* 27, 887-906.

12. Lagakos, S. W. (1976). A stochastic model for censored-survival data in the presence of an auxiliary variable. *Biometrics* 32, 551-559.
13. Lagakos, S. W. (1977). Using auxiliary variables for improved estimates of survival time. *Biometrics* 33, 399-404.
14. Lagakos, W. W., Sommer, C. J. and Zelen, M. (1978). Semi-Markov models for partially Censored data. *Biometrika* 65, 311-317.
15. Miller, R. (1981). *Survival Analysis*. Wiley, New York.
16. Peterson, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.* 72, 854-858.
17. Temkin, N. (1978). An analysis for transient states with application to tumor shrinkage. *Biometrics* 34, 571-580.
18. Voelkel, J. G. (1980). Multivariate counting processes and the probability of being in response functions. Ph. D. thesis, Department of Statistics, University of Wisconsin-Madison.
19. Voelkel, J. G. and Crowley, J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations. *Ann. Statist.* 12, 142-160.
20. Weiss, G. H. and Zelen, M. (1965). A semi-Markov model for clinical trials. *J. Appl. Prob.* 2, 269-285.
21. Tsai, W. Y. (1982). Bivariate survival time and censoring. Ph. D. thesis, Department of Statistics, University of Wisconsin-Madison.
22. Tsai, W. Y., Leurgans, S. and Crowley, J. (1986). Nonparametric estimation of a bivariate survival function in the presence of censoring. *Ann. Statist.* 14, 1351-1365.
23. Yeo, S. (1987). On estimation for a combined Markov and semi-Markov model with censoring, Ph. D. thesis, Department of Statistics, The Ohio State University.